BIOCHEMICAL STUDIES OF THE EUKARYOTIC RNA EXOSOME

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Matthew Alan Bratkowski

August 2012

BIOCHEMICAL STUDIES OF THE EUKARYOTIC RNA EXOSOME

Matthew Alan Bratkowski, Ph. D.

Cornell University 2012

The RNA exosome is a multi-subunit protein complex involved in RNA maturation, surveillance, and turnover of cellular RNA. It is composed of a nine subunit core that contains an RNA binding central channel and associates with the additional subunit Rrp44. Rrp44 contains a 3' to 5' processive exoribonuclease activity and an endoribonuclease activity in its N-terminal PIN domain. The exosome is present in both the nucleus and cytoplasm. In the nucleus, the exosome associates with a second nuclease, Rrp6, which is a distributive 3' to 5' exoribonuclease. Although a wealth of studies have helped to characterize the exosome, it is still not entirely clear how the exosome core functions in concert with these two nucleases to degrade RNA.

In this dissertation, I present structural and biochemical studies of both the Rrp44 and Rrp6 bound exosome. In particular, I have attempted to crystallize a sub-complex of Rrp44 bound to the exosome, and present data that suggest that structured RNA can access Rrp44 without being first channeled through the exosome. Despite this conclusion, the exosome core still appears to down-regulate both the exonuclease and endonuclease activities of Rrp44, possibly by causing the protein to assume a conformation that prohibits it from promiscuous RNA degradation. Work on the Rrp6 containing exosme was conducted in attempts of obtaining an EM reconstruction of this complex. Crystallization of an RNA and Rrp6 binding exosome associated

protein, Rrp47, was also attempted alone and in complex with Rrp6. Overall, the results that I present here will help to further characterize the function of the exosome in RNA mediated degradation.

BIOGRAPHICAL SKETCH

Matthew Bratkowski was born on July 25, 1983 in St. Louis, Missouri to Thomas and Gloria Bratkowski. He participated in Cub Scouts and Boy Scouts as a youth and received the rank of Eagle Scout in 2001. He attended Metro Academic and Classical Academy in St. Louis, where he was inducted into the National Honor Society, and graduated in 2002. In Fall 2002, he begin studies at the University of Missouri—Columbia. While he initially started college as an English major, after one year he switched to biology. Matthew realized that he enjoyed the challenge of the chemistry courses required for the biology major, and decided to add chemistry as a second major. He gained research experience in molecular biology in the laboratory of Miriam Golomb, where he was supported by the Life Sciences Undergraduate Research Award. He graduated in 2006 with a BS in Biological Sciences (with Honors, cum laude) and a BS in Chemistry (cum laude). In 2006, he was admitted to the Biochemistry, Molecular, and Cell Biology program at Cornell University and joined the lab of Ailong Ke. He later transferred to the Biophysics program at Cornell where he completed his dissertation.

ACKNOWLEDGMENTS

As long as I can remember, I have wanted to pursue a PhD degree, and I believe that this desire came from two sources. This first would be my father, who holds a PhD and is a college professor. Growing up, it always seems to me that he had the answer to any question that I asked. My second inspiration would be the fictional adventurer Dr. Indiana Jones, who gave class lectures before hunting for enchanted relics. Therefore, I always thought that obtaining a PhD would be an adventure, and it was.

I would first like to thank my parents, Tom and Gloria Bratkowski, for instilling upon me the values of an education. I would like to thank my two brothers, Mark and Tad Bratkowski (who both hold graduate degrees) for encouragement. I want to thank my girlfriend, Xin Luo (a fellow PhD student), for emotional support as well as for help in my scientific endeavors. I would also like to thank all of my fellow classmates for providing a much needed source of social interaction during graduate school. Finally, I would like to thank all of the present and past members of my laboratory: Ailong Ke, Fran Ding, Ray Lu, Ki Hyun Nam, Jason Griggs, and Ian Price. You all provided me with much guidance during my graduate career.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

α-^{32}P-CTP, alpha phosphate 32-labeled cytosine triphosphate

ATP, adenosine triphosphate

bp, base pair

CHES, *N*-Cyclohexyl-2-aminoethanesulfonic acid

cDNA, complementary DNA

CSD, cold shock domain

CTP, cytosine triphosphate

DMSO, dimethyl sulfoxide

DNA, deoxyribonucleic acid

DNase, deoxyribonuclease

DTT, dithiothreitol

EDTA, ethylenediaminetetraacetic acid

EGTA, ethylene glycol tetraacetic acid

EM, electron microscopy

EMSA, electrophoretic mobility shift assay

^{32}γP-ATP, gamma phosphate-32 labeled adenosine triphosphate

GTP, guanine triphosphate

HDV, Hepatitis Delta Virus

HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

IgG, immunoglobulin G

IPTG, Isopropyl β-D-1-thiogalactopyranoside

LB, Lysogeny broth

LIC, ligation independent cloning

MBP, maltose binding protein

mRNA, messenger ribonucleic acid

Ni-NTA, nickel-nitrilotriacetic acid

nt, nucleotide

OD_{600}, optical density at wavelength of 600 nanometers

OH, hydroxyl radical

PCR, polymerase chain reaction

PEG, polyethylene glycol

PIN, PilT N-terminus

PMSF, phenylmethanesulfonylfluoride

RNA, ribonucleic acid

RNase, ribonuclease

rpm, revolutions per minute

Rrp, ribosomal RNA processing

rRNA, ribosomal ribonucleic acid

SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis

snRNA, small nuclear ribonucleic acid

snoRNA, small nucleolar ribonucleic acid

TAP, tandem affinity purification

TEV, tobacco etch virus

Tris, tris(hydroxymethyl)aminomethane

tRNA, transfer ribonucleic acid

UTP, uridine triphosphate

UV, ultra violet

Chapter 1: Introduction to the RNA exosome

1.1. The Role of the Eukaryotic Exosome in RNA metabolism

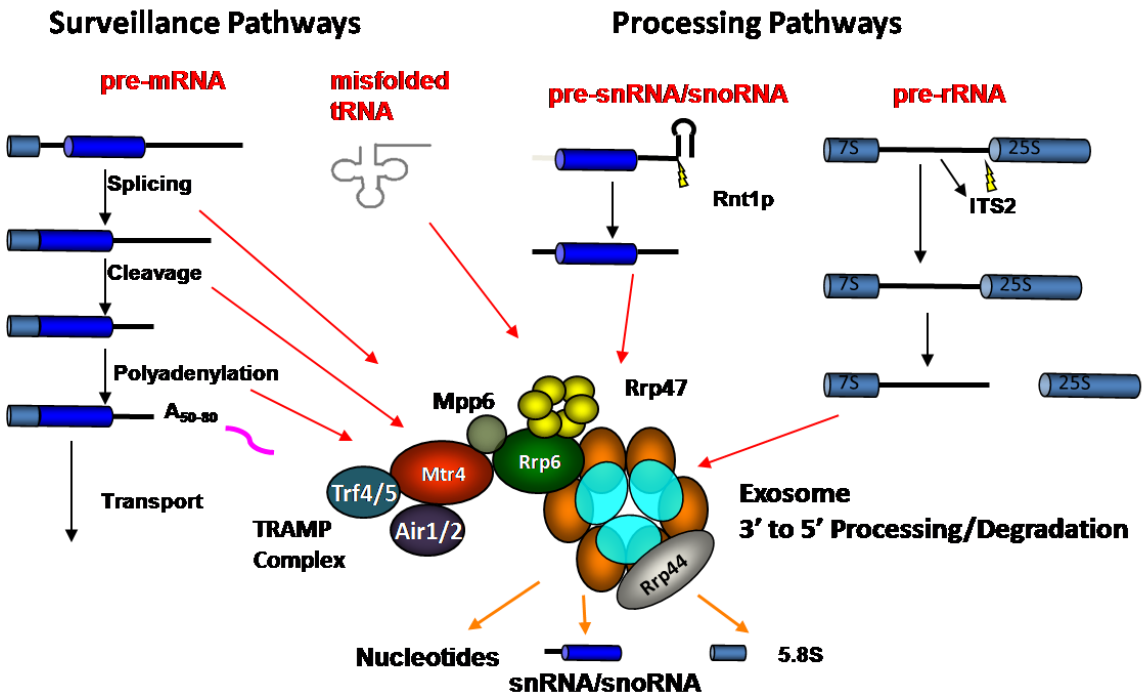
Virtually every RNA species in the cell is transcribed as a longer precursor that needs to be further processed before it is biologically active, a process known as RNA maturation. Unfortunately, defects arise during RNA transcription and processing, and aberrant RNAs must be rapidly degraded and rid from the cell by several different mechanisms collectively known as RNA surveillance. Once correctly processed RNAs perform their cellular functions and reach the end of their life-spans, they are degraded during RNA turnover. The aforementioned processes are controlled by either 5' to 3' or 3' to 5' exoribonucleases. The 3' to 5' RNA processing and degradation pathways are completed by the RNA exosome.

The exosome is a multi-subunit protein complex involved in RNA maturation, surveillance, and turnover (*1-7*). It degrades RNA starting at the 3' end of the RNA and degrades the RNA until it is completely degraded to 4-5 nucleotide end products in the RNA surveillance and turnover pathways, or until it cleaves off the precursor region of the premature RNA in the RNA maturation pathway (*8-10*). Since the exosome plays an essential role in numerous RNA metabolism pathways, it is not surprising that an intact exosome complex is necessary for cell viability: deletions of any of the core exosome subunits result in cell death in both yeast (*2*) and humans (*11, 12*).

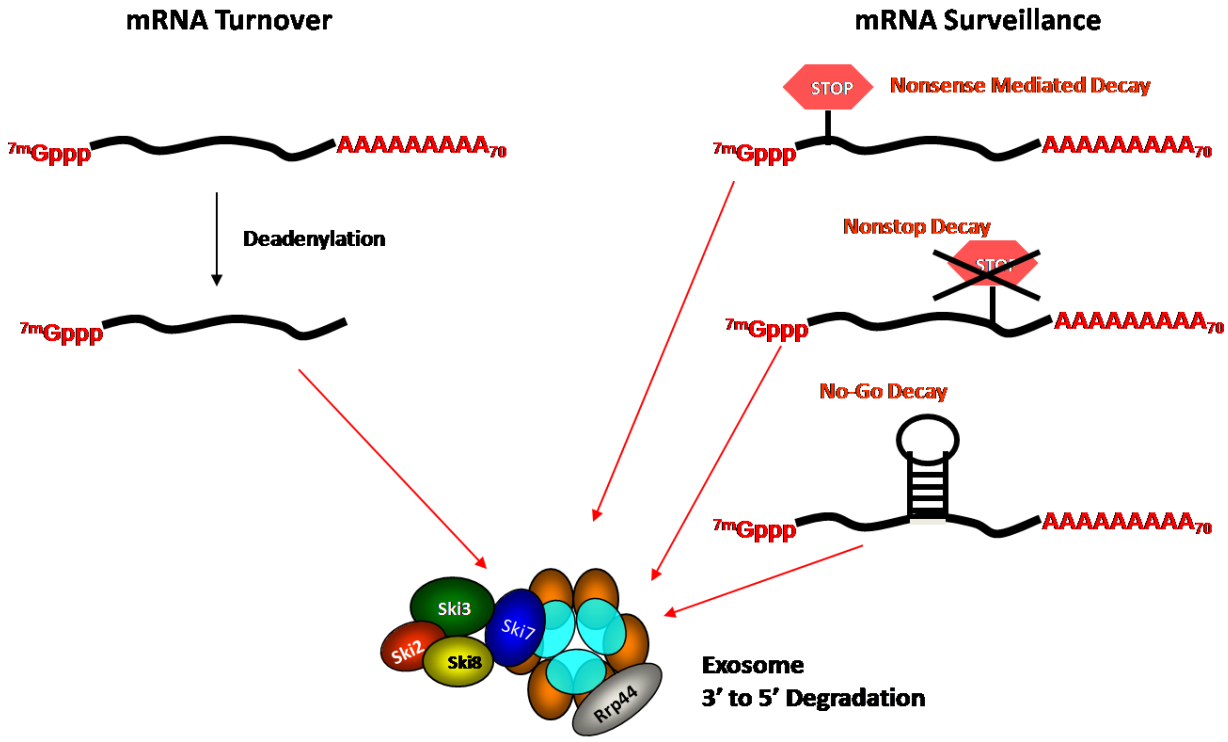
The eukaryotic exosome is a multi-subunit protein complex composed of a nine-subunit core that forms a double-layered ring (*13*). To degrade RNA, the exosome must associate with Rrp44, a 3' to 5' hydrolytic exoribonuclease *(12, 14)*. The exosome is present in both the nucleus and the cytoplasm, and functions in different RNA processing pathways in these

Figure 1-1. Role of the eukaryotic RNA Exosome. A) The nuclear exosome is involved in the surveillance of aberrantly processed mRNA and tRNA, and the 3' end maturation of snRNA, snoRNA, and rRNA. The exosome processes some snRNA and snoRNA by use of the nuclear specific nuclease Rrp6, which associates with the hexameric RNA binding protein Rrp47. In the nucleus, the TRAMP complex—which contains subunits with polyadenylase and helicase activities—helps to prepare substrates for degradation by the exosome. B) In the cytoplasm, the exosome is involved in mRNA turnover and surveillance. The cytoplasmic exosome associates with the SKI complex, which has a similar function to the nuclear TRAMP complex but contains Ski7, a putative ribosomal release factor.

A


Surveillance Pathways
Processing Pathways
pre-mRNA
misfolded tRNA
pre-snRNA/snoRNA
pre-rRNA
Splicing
Cleavage
Polyadenylation
A50-80
Transport
Rnt1p
7S
25S
ITS2
Mpp6
Rrp47
Mtr4
Rrp6
Trf4/5
Air1/2
TRAMP Complex
Exosome
3' to 5' Processing/Degradation
Rrp44
Nucleotides
snRNA/snoRNA
5.8S

B


mRNA Turnover
mRNA Surveillance
7mGppp
AAAAAAAAA70
Deadenylation
STOP
Nonsense Mediated Decay
Nonstop Decay
No-Go Decay
Ski3
Ski7
Ski2
Ski8
Rrp44
Exosome
3' to 5' Degradation

different compartments (Fig. 1-1). In the nucleus (Fig. 1-1A), the exosome is involved in RNA surveillance, degrading mRNA precursors that fail any mRNA processing steps such as splicing, cleavage, polyadenylation, or export out of the nucleus (*7*), as well as in the degradation of initiator methionine tRNA that is hypomethylated (*5*). It is involved in the 3' end maturation of a variety of non-coding RNAs such as small nuclear and small nucleolar (sn and sno) RNA (*1, 4*) (Fig. 1.1A). It is also involved in some steps of ribosomal RNA maturation, such as the degradation of the 3' end external transcribed spacer (*3*).

In the cytoplasm, the exosome is mainly involved in the turnover and surveillance of mRNA (Fig. 1-1B). During turnover, the exosome degrades mRNAs starting from their 3' end after they have first been deadenlyated. The exosome degrades aberrant mRNAs in a number of surveillance pathways. Surveillance pathways include Nonsense Mediated Decay, where the mRNA contains a premature stop codon (*15*); Non-stop Decay, where the mRNA is missing a stop codon (*9, 10*); and No-Go Decay, where the mRNA contains a strong secondary structure in the middle of its sequence that cannot be translated through (*16*).

In order to function in some of the aforementioned decay pathways, the exosome associates with additional proteins and accessory protein complexes in a compartment specific manner. In the nucleus, the exosome associates with the 3' to 5' exoribonuclease Rrp6 (Fig. 1.1A). Rrp6 appears to function in snRNA and snoRNA 3' end maturation, as well as some steps of rRNA maturation (*17*). Rrp6 also associates with the RNA binding protein Rrp47 in the processing of some of these RNA species. Additionally, the exosome associates with the TRAMP complex composed of the polyadenylase Trf4 (or Trf5), the zinc knuckle RNA binding protein Air1 (or Air2), and the helicase Mtr4 (Fig. 1A) (*18-20*). Trf4/5 polyadenylates RNA on its 3' end tail, which is thought to mark it as a substrate for the exosome, and Mtr4 assists to unwind the RNA

before degradation by the exosome (*18-20*). Air1 likely binds and positions the RNA while these processes occur.

In the cytoplasm, the exosome associates with the SKI complex formed by Ski2, Ski3, and Ski8 (Fig. 1.1B), which appears to play a role similar to the nuclear TRAMP complex, although it lacks a polyadenylation subunit (*8, 21*). In addition, the SKI complex contains the putative GTPase Ski7, which has been shown to bind to the exosome (*8-10*) and has a similar protein sequence as eukaryotic ribosome release factor (eRF3) (*22*). Therefore, it is hypothesized that Ski7 mimics eRF3 by binding to the ribosome and releasing aberrant mRNA from the ribosome to prevent it from being translated (*23*). The aberrant mRNA could then be subsequently shuttled to the exosome and degraded. While Ski7 has been implicated in the Non-Stop Decay surveillance pathway (*9, 10*), more research needs to be conducted to prove this hypothesis.

1.2. Structural and Functional Evolution of the Exosome

Exosome like structures are found in all three domains of life. All exosome like structures are composed of a six member ring containing subunits homologous to the bacterial RNase PH (Fig. 1-2A) (*24, 25*). RNase PH is hexameric complex of identical subunits, each of which contains phosphoroylytic 3' to 5' exoribonuclease activity: a molecule of phosphate is added to the 3' end of the RNA to catalyze degradation of the phosphodiester bond between two nucleotides. The main role of RNase PH appears to be the removal of nucleotides following the 3' end CCA tail of tRNA (*26*). Bacteria also contain another 3' to 5' phosphoroyltic exoribonuclease that is related to RNase PH, but more general in function, called the polynucleotide phosphorylase (PNPase). The PNPase contains three identical subunits. Each is composed of two central RNase PH domains that are attached to C-terminal KH and S1 domains (Fig. 1-2B) (*27, 28*). Though technically a trimer, the presence of dual RNase PH domains (1 and 2) in each of the

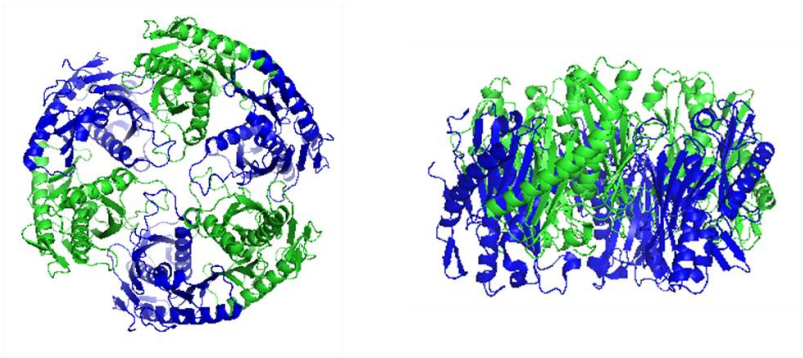
subunits of the PNPase allows it to fold into a similar hexameric shape as other exosome structures. In the PNPase, only RNase PH Domain 2 is catalytic active; thus, the complex contains only three phosphoroyltic active sites. In addition to its degradation activity, the PNPase can also catalyze the reverse reaction by acting as a polyadenosine polymerase to add nucleotides onto the 3' end of RNA. The PNPase is also present in eukaryotes in both chloroplasts and mitochondria, indicating that it retains a compartment specific role in eukaryotes (*29*).

The archaeal exosome is most similar in structure to the eukaryotic version. It is similar to its bacterial homologues, yet contains a few extra subunits. The archaeal exosome is composed of three heterodimers of RNase PH like subunits Rrp41 and Rrp42 that form a central ring, which is capped by three copies of the RNA binding protein Rrp4 or Csl4 (*30, 31*) to form a double layered ring structure (Fig. 1-2C). Rrp4 and Csl4 contain RNA binding domains similar to those found in the PNPase: both Rrp4 and Csl4 contain an S1 domain, and Rrp4 contains an additional KH domain. RNA is channeled through the central ring of the exosome into its active site. A continuous pathway of RNA progression through the channel is indicated by crystal structures of archaeal exosomes showing that one nucleotide binds near the opening of the channel or the "neck" region, while up to 6 nucleotides bind in the active site formed by subunits Rrp41 and Rrp42 (*32, 33*). In the active site, Rrp41 serves as the catalytic subunit and degrades the RNA using a 3' to 5' phosphoroyltic mechanism, while the Rrp42 subunit is inactive. Thus, similar to the PNPase, the archaeal exosome contains three active sites. Also like the PNPase, the archaeal exosome can add polyadenosine to the 3' end of RNA molecules (*30, 32, 34*).

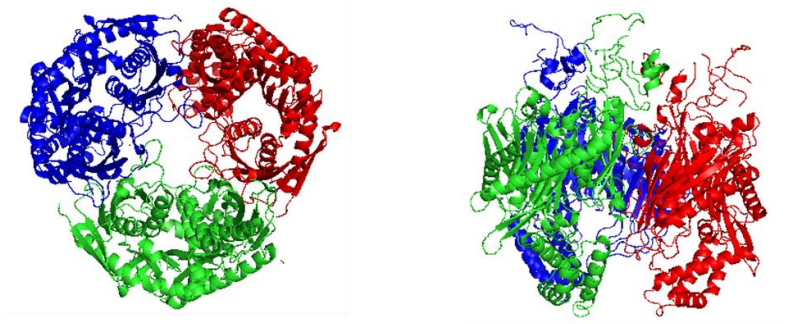
The eukaryotic RNA exosome core features an architecture similar to that of the archaeal exosome, but is composed of a wider diversity of subunits that comprise the double layered ring. It contains six RNase PH like channel subunits (Rrp41, Rrp45, Rrp42, Rrp43, Mtr3, and Rrp46), and three different RNA capping subunits (Rrp4, Rrp40, and Csl4) (*2, 13*). As shown by the structure of the human exosome (Fig. 1-2D), the eukaryotic exosome has an overall similar architecture to the archaeal exosome (*13*). Unlike the archaeal exosome, however, the channel of

Figure 1-2. Structural Evolution of the exosome. Structures are viewed on their side (left) and in a top to bottom orientation (right). A) RNase PH from *A. aeolicus* (pdb 1udn) is composed of six identical monomers. B) PNPase from *S. antibioticus* (pdb 1e3p) is a homotrimer. However, each subunit contains two RNase PH domains, so the structure mimics the RNase PH structure. Monomers contain additional C-terminal KH and S1 RNA binding domains that form loops near the entry of the channel. C) Archaeal exosome from *S. solfactaricus* (pdb 2jea). The RNase PH ring is composed of three heteordimers of Rrp41 and Rrp42. The ring is capped with three copies of Rrp4 (or Csl4, not shown) which contains RNA binding KH and S1 domains. Nucleotides found in the neck and processing chamber of the complex are displayed as spheres, and a probably pathway of RNA connecting these two binding sites is represented by a dashed line. D) Eukaryotic exosome from *H. sapiens* (pdb 2nn6). The RNase PH ring is formed by three heterodimers of Rrp41/Rrp45, Rrp42/Mtr3, and Rrp43/Rrp46. The RNA capping ring is formed by one copy of each of Rrp4, Rrp40, and Csl4. The figure was created using PYMOL software (*35*).

A. RNase PH
PH
PH

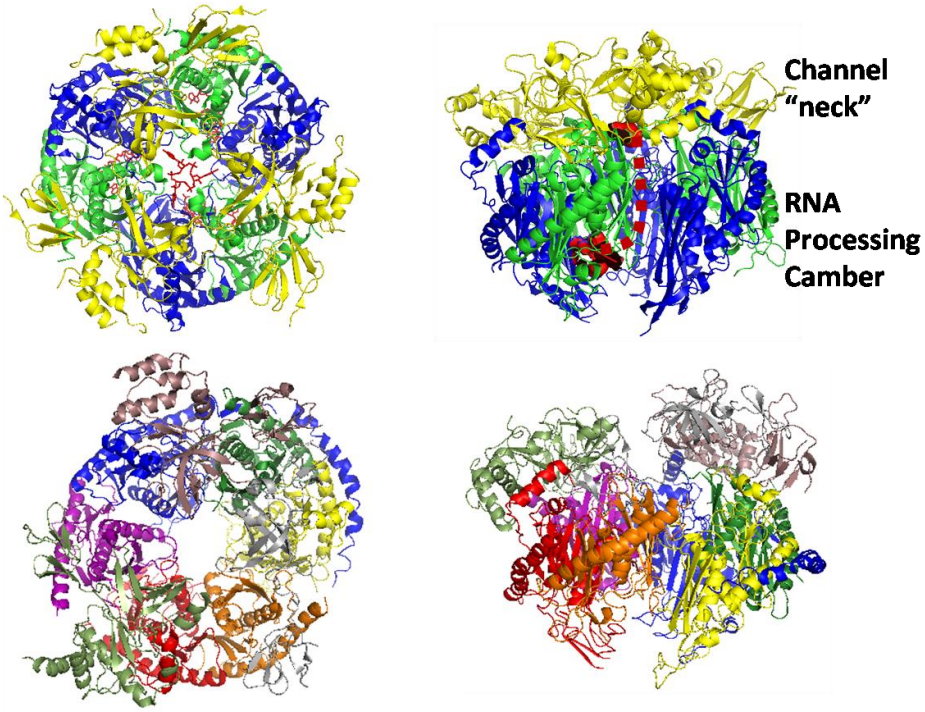


B. PNPase
PH1 and 2
PH 1 and 2
PH 1 and 2



C. Archael Exosome
Rrp41
Rrp42
Rrp4

D. Human Exosome
Rrp45
Rrp46
Rrp43
Rrp41
Mtr3
Rrp42
Csl4
Rrp40
Rrp4

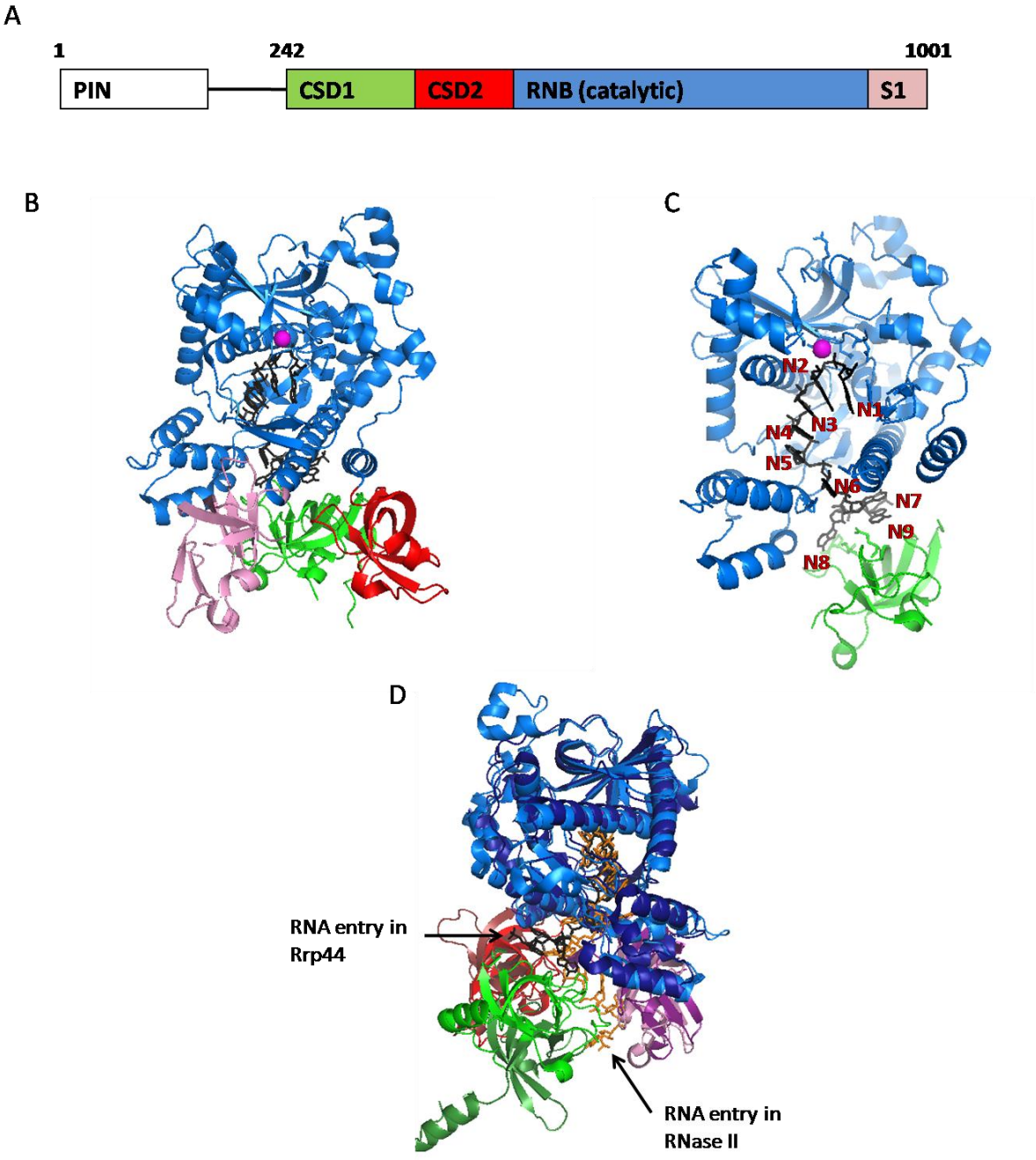


the eukaryotic exosome is catalytically inactive in both yeast (*14*) and humans (*13*), a trend that is likely conserved among eukaryotes, with some possible exceptions (*36*). Instead, activity is mainly provided by an additional subunit, the hydrolytic 3' to 5' exoribonuclease Rrp44 (also known as Dis3) (*14*). Rrp44 is a highly processive enzyme and appears to be bound to both nuclear and cytoplasmic exosomes in *S. cerevisiae* (*12, 14*). A second hydrolase, Rrp6, is found only associated with the nuclear exosome in yeast (*12, 37*). All core exosome subunits as well as Rrp44 and Rrp6 are conserved to an extent in higher eukaryotes, so the yeast exosome serves as a model eukaryotic exosome. In particular, the main catalytic subunit Rrp44 has been studied extensively in this organism.

1.3 Structure and function of the catalytic exosome subunit Rrp44

Rrp44 is composed of an N-terminal endoribonuclease (ENDO) PilT N-terminus (PIN) domain connected by a flexible linker to a C-terminal exoribonuclease (EXO) assembly that consists of two cold shock domains (CSD1 and CSD2), a large catalytic RNB domain, and a S1 domain (*38*) (Fig. 1-3A). PIN domains from proteins of various archae (*39-42*), bacteria (*43*), and humans (*44, 45*) contain a similar α/β/α fold. Some PIN domains contain endonuclease activity (*44*), but no structure has yet to be solved of a PIN domain bound to RNA. The structure of Rrp44ΔPIN (amino acids 242-1001) catalytically inactive mutant in complex with a single stranded polyadenine RNA revealed the topology of the body region of the protein and displayed its possible mechanism of RNA degradation (*38*). CSD1, CSD2, and S1 consist of a general oligosaccharide/oligonucleotide binding (OB) fold, which has five anti-parallel β strands that form a β-barrel (*46*). The RNB domain is mostly α helical in nature, and has an orientation very similar to the same domain in the *E. coli* exonuclease RNase II (*47, 48*). The RNB domain contains a channel that is ~10 Å wide, large enough to accommodate only single stranded RNA, that is lined with hydrophobic and positively charged residues leading to the active site of the protein (Fig. 1-3B-C) (*38*). Although the processing channel itself is narrow, its entrance is

Fig. 1-3. Structure of yeast Rrp44. A) Domain organization of the protein. B) Structure of *S. cerevisiae* Rrp44ΔPIN with nine polyadenine nucleotides bound in the active site (pdb 2vnu). Coloring corresponds to the domain organization depiction above. RNA is shown in black and the magnesium ion is in magenta. C) Cut-away view of the bound RNA with nucleotides labeled from the 3' end (N1) to the 5' end (N2). RNA makes most of its contacts in the RNB domain channel, and a few contacts with CSD1. D) Alignment of Rrp44 with *E. coli* RNase II containing bound RNA (pdb 2ix1). Rrp44 is presented in light shades while RNase II is depicted in dark shades. The bound RNA from the Rrp44 structure is shown as black sticks, while the RNA from the RNase II structure is in orange. Note that the secondary structure of the RNB domain is almost identical in the two proteins, but the RNA entry path varies greatly. This figure was created using PYMOL (*35*).


A
1
242
1001
PIN
CSD1
CSD2
RNB (catalytic)
S1
B
C
N2
N1
N3
N4
N5
N6
N7
N9
N8
D
RNA entry in
Rrp44
RNA entry in
RNase II

twice as wide and could accommodate duplex RNA. In the channel, 9 out of 13 nucleotides are bound and produce interpretable electron density. The five 3' most nucleotides bind in the catalytic center by means of electrostatic and base-stacking interactions with channel residues, and the nucleotides are threaded into the channel by binding between the CSD1 and RNB domains (*38*).

The complex structure of Rrp44 bound to a single stranded RNA substrate indicates that the RNA makes most of its contacts within this channel, with a few additional contacts to CSD1 near the entrance to the channel (Fig. 1-3C). Although this structure provides insight into how RNA binds in the RNB channel, the exact mechanism of RNA recruitment is still unclear. The structure is almost identical in domain organization to that of *E. coli* RNase II (which lacks a PIN domain), but RNase II directs the RNA towards its channel through contacts in the CSD2 and S1 domains instead (*47, 48*), which have yet to be implicated in RNA binding in Rrp44 (Fig 1-3D). Thus, it is unclear whether this different pathway represents divergent evolution of the two enzymes or whether the RNA binding site in the structure of Rrp44 is an artifact introduced by crystal lattice packing. If the RNA recruitment pathway proposed from the crystal structure is physiologically relevant, it does not appear that it functions in concert with the exosome in its crystallized form, since the 5' end of the RNA is pointed towards the solvent. Association of Rrp44 with the exosome, however, could result in a conformational change that would place the RNA entry site in line with the exit site of the exosome channel.

The EXO active site is located at the end of the channel in the RNB domain (*38*) and is used by Rrp44 to processively degrade one nucleotide at a time starting at the 3' end of the RNA. The active site contains density for one of two proposed magnesium ions, predicted to be involved in a two metal ion catalytic mechanism of degradation by Rrp44. The bound metal ion

is coordinated by D543, D552, and O3' of the phosphate from the second nucleotide, and the absent but proposed second magnesium ion is predicted to be coordinated by D549 and D551 (*38*). Since the D551N mutation inhibits exonuclease activity, it is likely that this mutation prevents coordination of the second magnesium ion, which results in its absence from the structure. The second metal ion is predicted to stabilize an activated water molecule and position it for attack on the scissile phosphate of the nucleotide at the 3' most end of the RNA (*47*). A particularly interesting facet of Rrp44 is its ability to process duplex RNA. The channel is only wide enough to accommodate single stranded RNA, yet Rrp44 can also degrade RNA duplexes with a long 3' overhang (*38*). The PIN domain of Rrp44 appears to be implicated in RNA unwinding by an unknown mechanism because Rrp44ΔPIN degrades duplexes with a short overhang much less efficiently than the full length version (*38*). The mechanism of RNA unwinding by Rrp44 is not fully understood because unlike RNA helicases, no ATP hydrolysis is involved to provide the energy required for unwinding. It is also not clear how this unwinding ability would function when Rrp44 is bound to the exosome core.

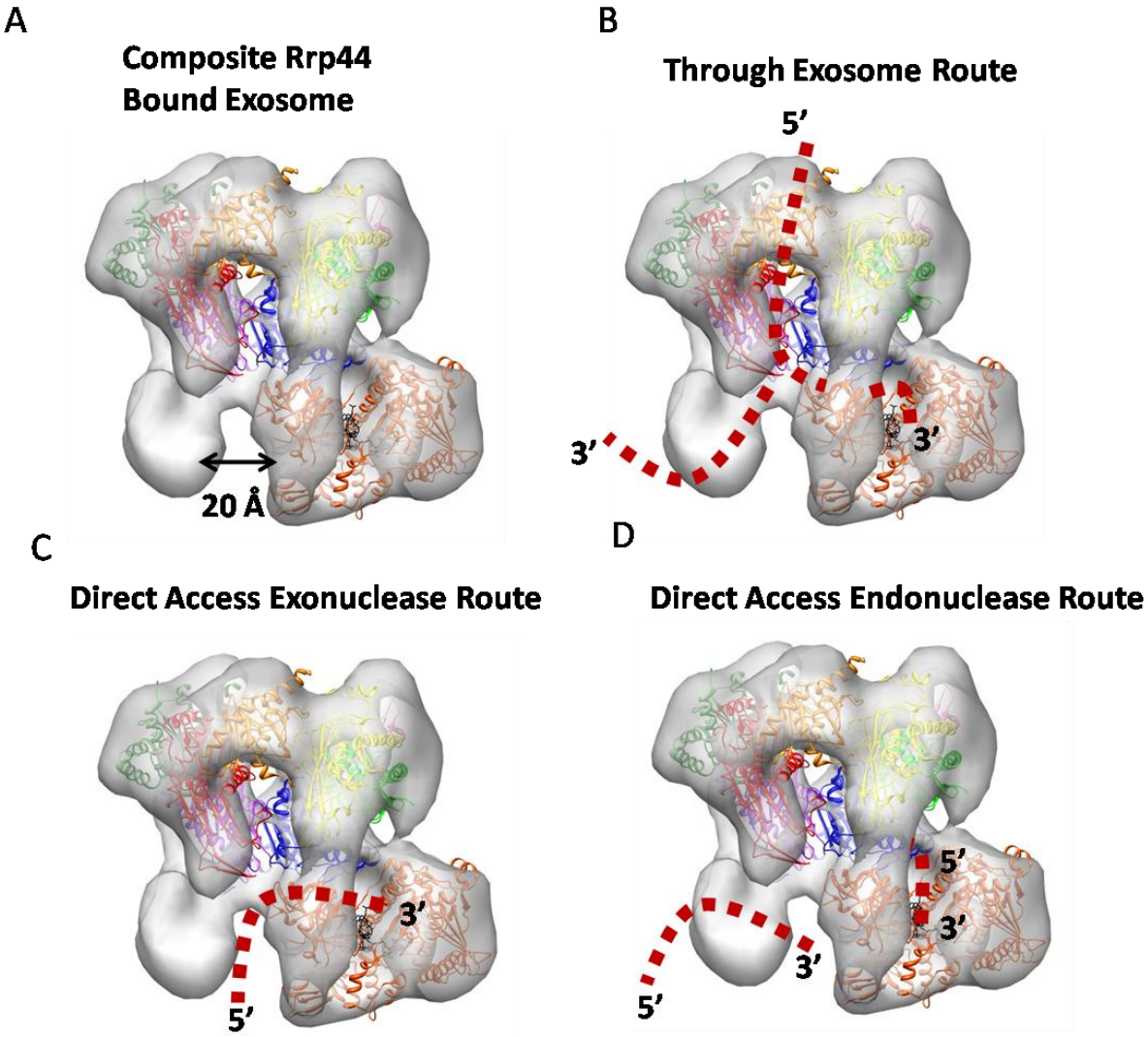
An ENDO active site resides in the PIN domain of Rrp44, which appears to be active on single stranded regions of RNA (*49-51*). While yeast cells that contain a mutation in either the EXO (D551N) or the ENDO (D171N) site are viable, mutation of both sites results in lethality (*49-51*). Interestingly, the two sites are optimally active at very different concentrations of metal ions. The EXO site is most active at low concentrations of Mg^{2+} (~100 uM) as indicated by *in vitro* studies, and inactive in high Mg^{2+} (5 mM) (*14*), while the ENDO site is active at high levels of Mn^{2+} (~3 mM) (*49-51*). It is a bit peculiar that the manganese concentration is much higher than physiological concentrations in cells. However, yeast genetic experiments indicate that both activities are used *in vivo*. It was reported that either active site can be used by Rrp44 to degrade

nonstop decay transcripts *in vivo*, yet degradation of normal mRNA during turnover was impeded by a mutation in only the exonuclease active site *(52),* and defects in 5.8S rRNA and snoRNA processing only occur as well with this mutation (*49-51*) . It has also been suggested that the PIN domain may assist in EXO mediated RNA unwinding (*38*). Likely, the PIN domain serves as an additional RNA binding platform that supports the substrate during degradation or directs it into the exonuclease active site of Rrp44. It is also possible that the two active sites work together to degrade certain substrates and a popular hypothesis is that the ENDO site creates new 3' ends on RNA substrates that are then degraded by the EXO site of Rrp44. An *in vitro* experiment supported this hypothesis by indicating that the endonuclease activity of Rrp44 cleaved single stranded loops in the 5' external transcribed spacer (5' ETS), a long excised region that is produced during rRNA maturation (*49*). However, since the two active sites are spatially distant, any synergy between them would need to be preceded by a conformational change in Rrp44 to bring the two sites into closer proximity, or would occur by the EXO and ENDO sites degrading different regions of the substrate either simultaneously or sequentially. Moreover, the mechanism of how the two sites would work together when Rrp44 is bound to the exosome is not yet known.

1.4. Association of Rrp44 with the Exosome Core

Given that the eukaryotic core exosome is inactive and that Rrp44 appears to achieve the catalytic activity for the complex, it is curious what the actual role of the exosome core is, and how it is involved in Rrp44 mediated RNA degradation. A negative stain EM reconstruction by Hongwei Wang in collaboration with the Ke group first revealed the architecture of the yeast

Figure 1-4. EM reconstruction of the yeast exosome and possible pathways of RNA recruitment. A) Reconstruction of the Rrp44 bound exosome (gray envelope, EM Databank ID 1438) with the crystal structure of the human exosome (pdb 2nn6) and yeast Rrp44ΔPIN (pdb 2vnu) docked in. Bound RNA is shown in black sticks. Docking was completed with UCSF Chimera (*54*). Note that the N-terminal PIN domain and the C-terminal catalytic region of Rrp44 appear to be located 20 Å apart. B) Possible "Through Exosome Route" reminiscent of the archaeal exosome. Note that the RNA could potentially interact with either the exo- or endo-nuclease sites after exosome channeling. C) The "Direct Access" exonuclease degradation pathway, in which the RNA would access the exonuclease site of Rrp44 in a manner similar to that in RNase II. D) "Direct Access" endonuclease pathway. This route could possibly be used to feed shortened RNAs into the exonuclease site during the degradation of more structured RNA.


A
Composite Rrp44
Bound Exosome
20 Å
B
Through Exosome Route
5′
3′
3′
C
Direct Access Exonuclease Route
3′
5′
D
Direct Access Endonuclease Route
5′
3′
3′
5′

exosome core and the yeast Rrp44 bound exosome (*53*). The exosome core reconstruction revealed a general ring and channel shape, and all subunits of the crystal structure of the human exosome could be docked into the reconstruction with a good fit, with the exception of Csl4, which appeared substoiciometric during purification.

The Rrp44 bound exosome reconstruction contained a ring structure in a similar conformation as that of the exosome core, but the large, two lobed density of ~140,000 $Å^3$ was attached to the bottom of the ring, corresponding to the 114 kDa Rrp44 protein (Fig. 1-4A) (*53*). About three-fourths of the density (~85 kDa) corresponded to the C-terminal region ("the body") of Rrp44, while the rest of the density (~28 kDa) resulted from the N-terminal PIN domain region ("the head") of Rrp44. The head and body regions of Rrp44 were separated by a distance of about 20 Å, and were connected by a flexible linker barely interpretable in the density. Upon analysis of the EM reconstruction, the head region of Rrp44 was determined to be a major site of the attachment of Rrp44 to the exosome, since this region buried a large surface area of 1,000 $Å^2$ into the exosome core subunit Rrp41. The body region of Rrp44 made most of its contacts to the exosome core through the Rrp45 subunit, burying a surface area of 1,800 $Å^2$ (*53*). A secondary contact was also established between the body of Rrp44 and the subunit Rrp43, burying a smaller surface area of 400 $Å^2$. Thus, Rrp44 appears to bind mainly to subunits Rrp41 and Rrp45 and is bound on the bottom side of the exosome core, situated underneath the channel.

Since Rrp44 is located on the exosome opposite the RNA binding cap and below the exit site of the channel, this places it in a position to receive RNA that is first funneled through the channel. The RNA could then pass into the exonuclease active site of Rrp44 for degradation, mimicking a mechanism similar to that of the archaeal exosome (Fig. 1-4B). Alternatively, the RNA could also potentially access the endonuclease site after exosome channeling, although this

may require a conformation change in Rrp44 since the PIN domain is mostly solvent exposed. A more recent Cryo-EM reconstruction of the Rrp44 bound exosome bound to a predominately single stranded RNA revealed extra electron density for RNA within the channel, yet the RNA was featureless and appeared loosely bound so it was not possible to clearly determine how it interacted with the channel (*55*). Rrp44 could also directly access the RNA substrate by means of either the solvent exposed pathway to its exonuclease site that was shown in the crystal structure (Fig. 1-4C), or its endonuclease PIN domain (Fig. 1-4D). The later pathway could potentially generate small RNA fragments that would then be feed into the exonuclease site for complete degradation. This pathway could be used by the exosome when degrading structured RNA, but its existence remains elusive.

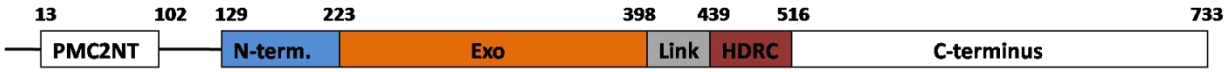
1.5. Role of Rrp6 and nuclear associated proteins in exosome mediated degradation

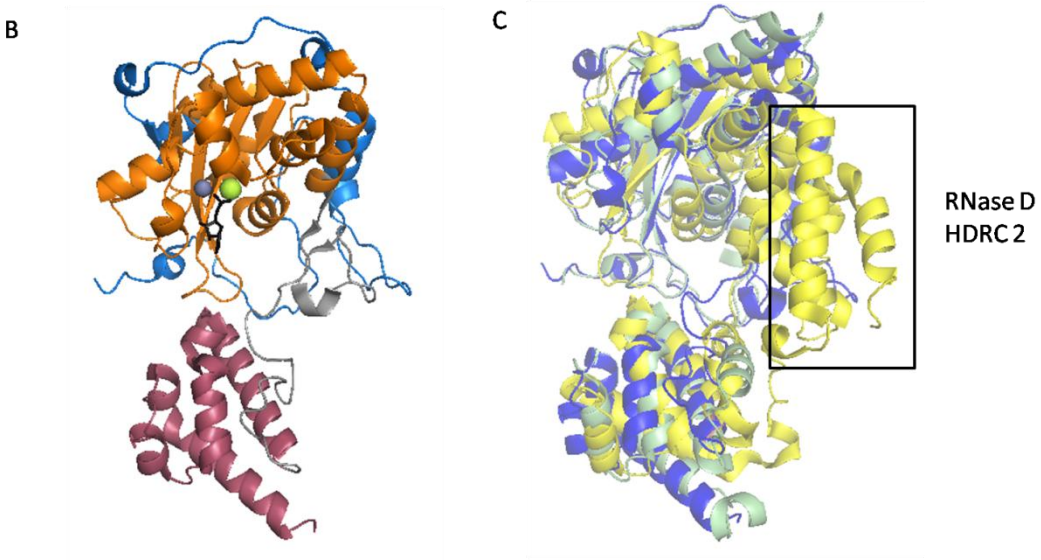
Rrp6 is 3' to 5' distributive hydrolytic exoribonuclease found in the nuclear exosome in *S. cerevisiae* that plays a prominent role in trimming snRNA, snoRNA, and portions of rRNA species into their mature forms (*56-58*). It contains an N-terminal domain that is predicted to be disordered, an exonuclease domain, a helicase and RNase D carboxy terminal (HDRC) domain, and a longer C-termius that is predicted to be unstructured (Fig. 1-5A). Crystal structures of the catalytic region of Rrp6 from yeast (amino acids 129-536, Fig. 1-5B) (*59*) and human (*60*) reveal a structure similar to that of RNase D (*61*) (Fig. 1-5C). However, RNase D includes an additional HDRC domain, has a shorter N-terminus, and lacks a disordered C-terminus. Rrp6 and RNase D are members of the DEDD family of RNA and DNA exonucleases, which contain nucleases such as RNase T (*62*) and the Klenow fragment of DNA polymerase I (*63*). Members

of this family use the DEDD residues to coordinate metal ions, which deprotonate a water molecule, creating a nucleophile that attacks the phosphodiester backbone of the RNA substrate (*64*). Specifically, RNase D and Rrp6 are part of a subset of DEDDy proteins that contain an additional tyrosine that coordinates the nucleophilic water. The distributive activity of Rrp6 is explained by the fact that its active is on the surface of the protein (*59*), preventing sequestering of the substrate which allows for processive degradation in Rrp44. Moreover, functional studies indicate that Rrp6 stalls upon encountering RNA stem loops (*13, 60*), suggesting its role in 3' end trimming but not complete degradation.

Although Rrp44 is an integral member of the exosome in yeast, the association of Rrp6 with the exosome and its role in RNA degradation is less clear. Pull-down assays indicated that the N-terminus of Rrp6 is required for binding to the exosome (*65*), but it is not known with which subunits it interacts in yeast. However, yeast two hybrid experiments conducted with human proteins indicated that Rrp6 interacts with Rrp41, Rrp43, Rrp46, and Mtr3, yet the location of Rrp6 on the exosome is unknown (*66*). Rrp6 is assumed to be positioned on the top of the exosome ring distal to Rrp44, and near or bound to the RNA cap, based on a negative stain EM structure of the *Leishmania tarentolae* exosome (*67*). However, the rather poor resolution of the reconstruction prevents unambiguous assignment of the Rrp6 binding site and its orientation, and it is unknown whether such an association is conserved in higher eukaryotes. A current mechanistic model indicates that Rrp6 processes structured RNA that does not enter the exosome channel, yet this hypothesis does not explain why it needs to be attached to the exosome to perform these processes. Also, despite its role in maturation of RNA, deletion of Rrp6 from *S. cerevisiae* does not result in lethality (*17*), which results by deleting any of the other exosome subunits (*2*). Moreover, functional exosomes can be isolated from Rrp6Δ yeast strains,

Figure 1-5. Structure of the catalytic region of yeast Rrp6. A) Domain organization of *S. cerevisiae* Rrp6. The numbering scheme corresponds to the amino acid length. Domains not found in the crystallized fragment are colored in white. B) Structure of the catalytic region (amino acids 129 – 536) of yeast Rrp6 bound to metal ions and UMP (pdb 2hbm). Domain coloring corresponds to the diagram above. Bound UMP is colored in black. A zinc ion near the active site is in grey and a manganese ion is in lime. C) Alignment of *S. cerevisiae* Rrp6 catalytic region (dark blue, pdb 2hbm) with the catalytic region of human Rrp6 (light green, pdb 3SAH) and *E. coli* RNase D (yellow, pdb 1YT3). The additional HDRC domain in RNase D is boxed. Figures B and C were created with PYMOL software (*35*).


13
102
129
223
398
439
516
733
PMC2NT
N-term.
Exo
Link
HDRC
C-terminus


B
C
RNase D
HDRC 2

suggesting that it is not necessary for the structural integrity of the exosome. Thus, it is possible that Rrp44 can compensate for some of the functions of Rrp6.

Another protein which plays a key role in nuclear exosome processing is Rrp47. Rrp47 is a basic protein that binds to the N-terminal PMC2NT domain of Rrp6, and it appears to function in the processing of snRNA and snoRNA, since deletion of Rrp47 results an accumulation of precursor RNAs which are similar to those occurring in Rrp6 deletion strains (*56-58*). Rrp47 possess no catalytic activity but forms a hexamer *in vitro* that can bind tightly to structured RNA and Rrp6 simultaneously (*57, 58*); interestingly, Rrp47 also binds double stranded DNA (*57*). It is hypothesized that hexamerization of Rrp47 may thus create an RNA binding channel that deposits the substrate into the active site of Rrp6, although no structural studies have yet to confirm this assumption. It appears that one function of Rrp47 is to bind to small nucleolar ribonucleoprotein (snoRNP) complexes to guide the processing of the 3' tails of C/D box snoRNAs—which are involved in rRNA methylation—by Rrp6 (*58*). Theories speculate that Rrp47 may serve as either an RNA binding platform for Rrp6 or to help regulate its activity. A second protein known as Mpp6 also associates with Rrp6 (*68, 69*), and may serve a similar role as Rrp47. It is likely that each protein is specific for certain substrates.

1.6 The role of Ski7 in cytoplasmic mRNA decay

Ski7 stably associates with the cytoplasmic exosome and is retained during TAP purifications of the complex. Ski7 has been suggested to bridge interactions between the exosome and the SKI complex during cytoplasmic exosome mediated degradation (*23*). The N-terminus of Ski7 is required for proper degradation of normal mRNAs, while its C-terminus is dispensable for this

function (*70*). Ski7 has also been implicated in the non-stop decay RNA surveillance pathway, where transcripts lacking a stop codon are rapidly degraded by the exosome, beginning with their 3' polyadenine tail (*9, 10*). Separate N-terminal portions of Ski7 bind to both the exosome and the Ski complex (Fig. 1-6A), and Ski7 is hypothesized to tether the complexes together to allow RNA to be unwound by Ski2 before being degraded by the exosome (*70, 71*). Ski7 is thought to bind to the cap proteins of the exosome since it co-purifies with Csl4 and Rrp4 (*9*), however, definitive experiments have yet to be completed to prove this hypothesis. In particular, the zinc-ribbon domain of Csl4 may serve as a primary binding site since it is essential for mRNA decay, but not decay of other species of RNA (*50*).

Although no structural information exists for Ski7, it is predicted to have GTPase activity based on the similarity of its C-terminal region (Ski7-c) to structures of elongation factor 1A (EF1A) and eukaryotic release factor 3 (eRF3). Crystal structures of EF1A from *S. cerevisiae* (Fig. 1-6B) (*72-74*) and the C-terminal region of eRF3 (eRF3-c) from *S. pombe* (Fig. 1-6C) (*75*) have a similar architecture. Both structures have a Domain 1 that contains a GKT/S P-loop that coordinates an Mg^{2+} ion (*73*), and binds the phosphate moiety of GTP/GDP as well as an NKxD motiff that binds the nucleotide ring of guanine. Domain 1 also contains the Switch I and Switch II helices that are responsible for the exchange of GTP for GDP (*72, 75*). Although the GTP/GDP binding sequence motifs are conserved in Ski7-c, the rest of Ski7-c does not align well with EF1A or eRF3 and contains several insertions (*76*). Furthermore, these regions are common to GTPases and do not necessarily suggest a function conserved beyond basic GTP hydrolysis. For instance, although eEF1A and eRF3c have a similar architecture, they have disparate roles in translation. During elongation, EF1A binds GTP and brings aminoacylated tRNA (aatRNA) (*77*) to the A site of the 80S ribosome, where the aatRNA binds to its specific

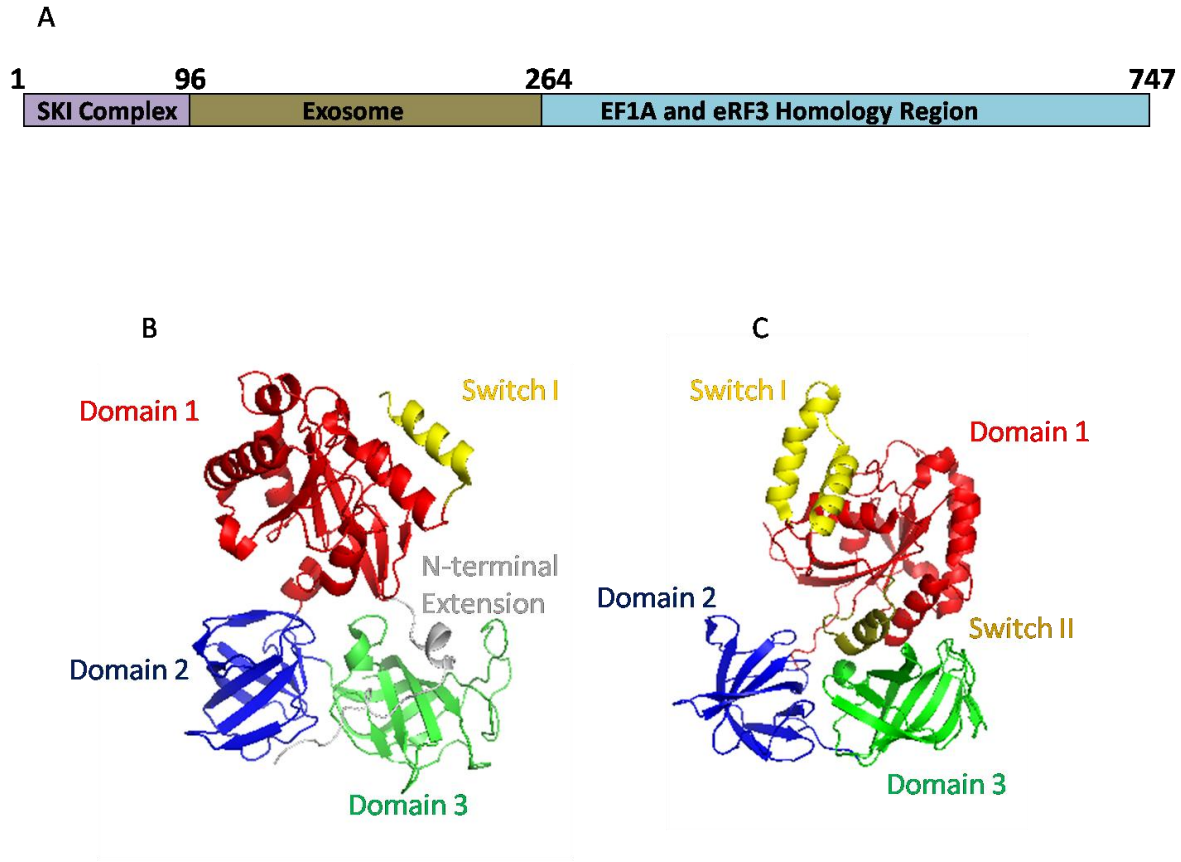
codon of the translating mRNA. On the other hand, eRF3 functions during termination by mimicking the structure of a tRNA to and recognizes all three stop codons (*78*). The GTPase activity of eRF3 is stimulated by binding to release factor 1 (eRF1) and the ribosome, and GTP hydrolysis drives peptide release from the ribosome.

Given some sequence similarity between eRF3 and Ski7 and its necessity in Non-Stop decay, it is an intriguing hypothesis that Ski7 is involved in peptide release in a stop codon independent manner. Stop codon independent termination of structured mRNA does occur in the No-Go pathway and requires Hbs1 and Dom34, eRF3 and eRF1 paralogues, respectively (*16*), which appear to function in a similar manner to cause peptide release. If Ski7 functioned in a similar manner, it would be expected to bind to a protein resembling eRF1 to cause polypeptide release, but no such binding partner has been identified. It is also possible that Sk7 has no direct role in termination but instead serves mainly as an adaptor to bring the exosome to the ribosome to allow for rapid degradation of mRNA after termination has occurred. A thorough investigation of the biochemical properties of Ski7 and its relation to the exosome must be conducted to clarify its role in Non-Stop decay.

1.7. Conservation of the exosome in eukaryotes

The exosome has been studied extensively in budding yeast and more recently in humans, and the core exosome appears to be for the most part conserved between these two species. One main exception is that subunit Rrp45 contains a long C-terminal extension in humans (*13*) that is absent in the *S. cerevisiae* exosome. This region wraps around the solvent exposed surface of subunits Rrp46 and Rrp43 (*13*), and could possibly provide additional crystal contacts that

Figure 1-6. Binding region of Ski7 and possible structural homology. A) Schematic diagram of amino acid regions of Ski7 shown to bind to the SKI complex (lavender) and the exosome (tan), as well as the region homologous to EF1A and eRF3 (light blue). The amino acid number is given above the diagram. B) Crystal structure of the C-terminal region of *S. pombe* eRF3 (pdb 1r5b). C) Crystal structure of the *S. cerevisiae* EF1A (modified from the EF1A/EF1Bα complex, pdb 1f60). Figures B and C were created using PYMOL software (*35*).


A
1
96
264
747
SKI Complex
Exosome
EF1A and eRF3 Homology Region
B
Domain 1
Switch I
N-terminal
Extension
Domain 2
Domain 3
C
Switch I
Domain 1
Domain 2
Switch II
Domain 3

allowed the human exosome to crystallize, while the yeast version is thus far recalcitrant to crystallization. Regardless, deleting this region from the crystal structure and docking the human exosome into the yeast exosome EM envelope results in a good fit (Fig. 1-4A), indicating that the yeast core exosome likely forms a similar structure.

While the core exosome subunits appear to be consistent in yeast and humans, some differences appear to occur in other eukaryotes. A few organisms lack particular subunits; for instance *D. melanogaster* exosomes lack subunit Rrp43. While the exosome core is inactive in yeast and humans, *A. thaliana* Rrp4 contains phosphoroyltic RNase activity (*79*) and rice Rrp46 forms a homodimer and contains both phoshoroyltic RNase and hydrolytic DNase activity (*80*). Therefore, it appears that in some eukaryotes the structure and function of the exosome has diverged.

The association of higher eukaryotic exosomes with Rrp44 differs between yeast and humans. While the yeast exosome binds tightly to Rrp44 during Tandem Affinity Purifications (TAP), Rrp44 was initially not retained in human exosome purifications (*81*), due to salt sensitive dissociation of Rrp44. The association of human Rrp44 with the exosome is also complicated by the fact that higher eukaryotes such as humans, mice, and frogs contain genes for three Rrp44 like proteins known as (using the alternative Dis3 nomenclature) hDis3, hDis3-like exonuclease 1 (hDis3L1), and hDis-3 like exonuclease 2 (hDis3L2). Both hDis3 and hDis3L1 contain regions predicted to form PIN domains, and thus could potentially interact with the exosome, while hDis3L2 has no PIN domain and likely does not interact with the exosome. Subsequent TAP purifications conducted in the presence of a low NaCl concentrations reveal that both hDis3 and hDis3L1 do associate with the exosome (*82, 83*). While both isoforms display exonuclease activity and contain a structural PIN domain (*82, 83*), only hDis3 has

endonuclease activity, (*82*) and can complement Dis3 depleted yeast cells yet hDis3L1 cannot. Furthermore, the two isoforms localize to different cellular compartments in humans. While hDid3 localizes to the cytoplasm and the nucleus (specifically the nucleoplasm), hDis3L1 is restricted to the cytoplasm (*82*). Although hDis3L2 lacks a PIN domain and presumably does not bind to the exosome or possess endonuclease activity, it is an active exonuclease and localizes to the cytoplasm (*84*). Therefore, the pattern of Rrp44-mediated exosome degradation appears more complex in human as opposed to yeast cells. Current information indicates that each Rrp44 isoform could have a compartment specific role in humans.

The human exosome also contains an analogue of Rrp6 known as PM-SCL-100 or hRrp6 that is known to be involved in nuclear RNA processing. Unlike in the yeast exosome, however, hRrp6 has been reported to localize to both the nucleus and cytoplasm (*11*, *15*). Structural comparison between the catalytic domains of the yeast and human Rrp6 indicate that the human version has a more exposed active site (Fig. 1-5C), and RNase assays indicate that this provides it with a greater ability to degrade structured RNA (*60*). Similar to the yeast exosome, hRrp6 is known to associate with human Rrp47 (C1D), and also human Mpp6 (MPHOSPH6) (*85*). Furthermore, it associates with the human nuclear exosome targeting (NEXT) complex composed of members of the TRAMP complex such as hMtr4, the zinc knuckle protein ZCCHC7 (an Air2 homologue), and hTrf4-2 (a Trf4 homologue) (*86*). Likely, compartment specific proteins allow for a different selection of RNA substrates in different compartments.

In contrast to Rrp44 and Rrp6, Ski7 has not been identified in humans. Although a putative Ski7 homology protein has been identified in *Schizosacchromycces pombe*, in general Ski7 is poorly conserved even within yeast species (*76*). Therefore, either functional—but not necessarily structural—homologues of Ski7 have yet to be identified in other species, or Ski7 is

restricted to a small class of yeast species.

1.8. Role of the exosome in human pathologies

Subunits of the human exosome were initially identified to be involved in human pathologies well before their role in RNA degradation was known. The human exosome is a well-established target of autoimmunity in the connective tissue disease polymyositis/scleroderma-overlap syndrome (PM/SCL) (*87-90*), and in fact this is the origin of the alternative names of several human subunits such as Rrp6 (PM-SCI-100) and Rrp45 (PM-SCL-75). PM/SCL is an autoimmune disorder: autoantibodies target a patient's own proteins, destroying their function. It is characterized by weakening of the muscles, impairing the patient's normal mechanical ability. In PM/SCL, autoantibodies frequently target hRrp4, hRrp6, hRrp45, and hRrp47, which are listed in the order of decreasing percentage of occurrence of autoantibodies found in the patient's serum (*91*). It is therefore likely that PM/SCL results from impaired RNA metabolism that results from the targeted destruction of exosome subunits. However, despite the advances made in exosome research over the past decade, little evidence exists as to why the exosome is targeted in PM/SCL or how this targeting results in muscle deterioration. A better mechanistic understanding of the exosome may thus shed insight into the molecular consequences of the disease on patients and possible ways to overcome them.

The exosome has also been implicated in cancer. Auto-antibodies attack the exosome subunit hRrp46 in patients suffering from chronic myelogenous leukemia (*92*), a cancer that causes the uncontrolled production of myeloid cells and represents about 20% of all adult leukemias (*93*). Although hRrp46 is the only subunit yet implicated in this disease, more

comprehensive studies may implicate other subunits as well. The exosome is also a secondary target of the chemotherapy drug 5-fluorouracil, which is used to treat colorectal and pancreatic cancers. Although the main target of this pyrimidine analogue is thymidylate synthase—resulting in inhibited thymine production—it has been found to inhibit the nuclease activity of Rrp6 in both yeast (*94*) and humans (*95*). Treatment with the drug results in the accumulation of precursor rRNA due to the inhibited processing ability of Rrp6 (*94*). Thus, Rrp6 could serve as a possible target of new cancer therapeutics.

Although the exosome is the target of autoimmune disease, it is essentially involved in protecting organisms from genetic and possibly pathogenic damage by degrading nonsense transcripts which cause about one third of all inherited human diseases. Further characterization of the role of the exosome in human nonsense mediated decay may thus be useful for determining how some nonsense transcripts fail to be degraded and are translated into aberrant proteins that cause disease.

1.9. Conclusions and areas of future study

Over the last thirteen years, a wealth of research has been conducted to characterize the structure and functional roles of the exosome. Structures of exosome like complexes have been determined in all three domains of life and maintain a similar multi-subunit ring. In particular, the crystal structure of the human exosome revealed how eukaryotic exosomes are composed. Despite the reoccurrence of the exosome core like structures throughout evolution, a key finding revealed that the eurkaryotic exosome core is inactive in most eukaryotes, except when associated with either Rrp44 or Rrp6 nucleases. Work on these two nucleases has provided us

with structures of the catalytic regions of both of them, as well as likely mechanisms of how they function to degrade RNA. Moreover, recent work has indicated how the human versions of these enzymes relate to their more studied yeast counterparts. Specifically, three isoforms of Rrp44 exist in human cells, but only two associate with the exosome, and human Rrp6 has a structure similar to yeast but has a more open active site to allow for less restricted RNA access. These findings provide us with a clearer understanding of how the exosome functions during RNA degradation.

The fact that the core exosome itself is inactive in eukaryotes, however, remains a curiosity. It is interesting to ponder whether the exosome core retains a functional role, or instead serves mainly as a scaffold for the attachment of nuclease subunits. While the location of where Rrp6 binds to the core has yet to be determined, it is generally accepted that Rrp6 degrades RNA without making use of the exosome channel, and it has even been suggested that it can function in a similar manner without even binding to the exosome (*65*). In contrast, EM reconstructions of the Rrp44 bound exosome clearly place Rrp44 on the underside of exosome channel (*53, 55*). While the eukaryotic core exosome has been suggested to function in a manner similar to its archaeal cousin during Rrp44 mediated RNA degradation, it largely remains to be determined whether RNA can access Rrp44 without being first being channeled through the core. This is an intriguing hypothesis, given that Rrp44 contains endonuclease activity in its PIN domain, a region that appears inaccessible to RNA traveling through the exosome core (*53*).

The main goals of our work are to more holistically characterize the structure of the eukaryotic exosome and the process of how it degrades RNA. We begin with structural and functional studies of the Rrp44 bound exosome, with a goal of testing the possibility of the "direct access" pathway of RNA recruitment to Rrp44. We characterize the association of the

nuclear exosome with Rrp6, Rrp47, and various other nuclear cofactors in attempts of obtaining detailed information on the structural associations of these proteins with the exosome. The work presented here as well as continued studies will help to define the possibly mechanisms of RNA processing by the eukaryotic exosomes.

References

1. Allmang, C., Kufel, J., Chanfreau, G., Mitchell, P., Petfalski, E., and Tollervey, D. (1999) Functions of the exosome in rRNA, snoRNA and snRNA synthesis, *The EMBO journal 18*, 5399-5410.
2. Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M., and Tollervey, D. (1997) The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-->5' exoribonucleases, *Cell 91*, 457-466.
3. Allmang, C., Mitchell, P., Petfalski, E., and Tollervey, D. (2000) Degradation of ribosomal RNA precursors by the exosome, *Nucleic acids research 28*, 1684-1691.
4. van Hoof, A., Lennertz, P., and Parker, R. (2000) Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs, *Molecular and cellular biology 20*, 441-452.
5. Kadaba, S., Krueger, A., Trice, T., Krecic, A. M., Hinnebusch, A. G., and Anderson, J. (2004) Nuclear surveillance and degradation of hypomodified initiator tRNAMet in S. cerevisiae, *Genes & development 18*, 1227-1240.
6. Schneider, C., Anderson, J. T., and Tollervey, D. (2007) The exosome subunit Rrp44 plays a direct role in RNA substrate recognition, *Molecular cell 27*, 324-331.
7. Bousquet-Antonelli, C., Presutti, C., and Tollervey, D. (2000) Identification of a regulated pathway for nuclear pre-mRNA turnover, *Cell 102*, 765-775.
8. Anderson, J. S., and Parker, R. P. (1998) The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex, *The EMBO journal 17*, 1497-1506.
9. van Hoof, A., Frischmeyer, P. A., Dietz, H. C., and Parker, R. (2002) Exosome-mediated recognition and degradation of mRNAs lacking a termination codon, *Science 295*, 2262-2264.
10. Frischmeyer, P. A., van Hoof, A., O'Donnell, K., Guerrerio, A. L., Parker, R., and Dietz, H. C. (2002) An mRNA surveillance mechanism that eliminates transcripts lacking termination codons, *Science 295*, 2258-2261.
11. van Dijk, E. L., Schilders, G., and Pruijn, G. J. (2007) Human cell growth requires a functional cytoplasmic exosome, which is involved in various mRNA decay pathways, *RNA 13*, 1027-1035.
12. Allmang, C., Petfalski, E., Podtelejnikov, A., Mann, M., Tollervey, D., and Mitchell, P. (1999) The yeast exosome and human PM-Scl are related complexes of 3' --> 5' exonucleases, *Genes & development 13*, 2148-2158.
13. Liu, Q., Greimann, J. C., and Lima, C. D. (2006) Reconstitution, activities, and structure of the eukaryotic RNA exosome, *Cell 127*, 1223-1237.
14. Dziembowski, A., Lorentzen, E., Conti, E., and Seraphin, B. (2007) A single subunit, Dis3, is essentially responsible for yeast exosome core activity, *Nature structural & molecular biology 14*, 15-22.
15. Lejeune, F., Li, X., and Maquat, L. E. (2003) Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylating, and exonucleolytic activities, *Molecular cell 12*, 675-687.
16. Doma, M. K., and Parker, R. (2006) Endonucleolytic cleavage of eukaryotic mRNAs

with stalls in translation elongation, *Nature 440*, 561-564.
17. Briggs, M. W., Burkard, K. T., and Butler, J. S. (1998) Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential for efficient 5.8 S rRNA 3' end formation, *The Journal of biological chemistry 273*, 13255-13263.
18. Kadaba, S., Wang, X., and Anderson, J. T. (2006) Nuclear RNA surveillance in Saccharomyces cerevisiae: Trf4p-dependent polyadenylation of nascent hypomethylated tRNA and an aberrant form of 5S rRNA, *RNA 12*, 508-521.
19. LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A., and Tollervey, D. (2005) RNA degradation by the exosome is promoted by a nuclear polyadenylation complex, *Cell 121*, 713-724.
20. Vanacova, S., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., Langen, H., Keith, G., and Keller, W. (2005) A new yeast poly(A) polymerase complex involved in RNA quality control, *PLoS biology 3*, e189.
21. Brown, J. T., Bai, X., and Johnson, A. W. (2000) The yeast antiviral proteins Ski2p, Ski3p, and Ski8p exist as a complex in vivo, *RNA 6*, 449-457.
22. Benard, L., Carroll, K., Valle, R. C., Masison, D. C., and Wickner, R. B. (1999) The ski7 antiviral protein is an EF1-alpha homolog that blocks expression of non-Poly(A) mRNA in Saccharomyces cerevisiae, *Journal of virology 73*, 2893-2900.
23. Maquat, L. E. (2002) Molecular biology. Skiing toward nonstop mRNA decay, *Science 295*, 2221-2222.
24. Harlow, L. S., Kadziola, A., Jensen, K. F., and Larsen, S. (2004) Crystal structure of the phosphorolytic exoribonuclease RNase PH from Bacillus subtilis and implications for its quaternary structure and tRNA binding, *Protein science : a publication of the Protein Society 13*, 668-677.
25. Ishii, R., Nureki, O., and Yokoyama, S. (2003) Crystal structure of the tRNA processing enzyme RNase PH from Aquifex aeolicus, *The Journal of biological chemistry 278*, 32397-32404.
26. Li, Z., and Deutscher, M. P. (1996) Maturation pathways for E. coli tRNA precursors: a random multienzyme process in vivo, *Cell 86*, 503-512.
27. Shi, Z., Yang, W. Z., Lin-Chao, S., Chak, K. F., and Yuan, H. S. (2008) Crystal structure of Escherichia coli PNPase: central channel residues are involved in processive RNA degradation, *RNA 14*, 2361-2371.
28. Symmons, M. F., Jones, G. H., and Luisi, B. F. (2000) A duplicated fold is the structural basis for polynucleotide phosphorylase catalytic activity, processivity, and regulation, *Structure 8*, 1215-1226.
29. Schuster, G., and Stern, D. (2009) RNA polyadenylation and decay in mitochondria and chloroplasts, *Progress in molecular biology and translational science 85*, 393-422.
30. Buttner, K., Wenig, K., and Hopfner, K. P. (2005) Structural framework for the mechanism of archaeal exosomes in RNA processing, *Molecular cell 20*, 461-471.
31. Lorentzen, E., and Conti, E. (2005) Structural basis of 3' end RNA recognition and exoribonucleolytic cleavage by an exosome RNase PH core, *Molecular cell 20*, 473-481.
32. Lorentzen, E., Walter, P., Fribourg, S., Evguenieva-Hackenberg, E., Klug, G., and Conti, E. (2005) The archaeal exosome core is a hexameric ring structure with three catalytic subunits, *Nature structural & molecular biology 12*, 575-581.
33. Navarro, M. V., Oliveira, C. C., Zanchin, N. I., and Guimaraes, B. G. (2008) Insights into the mechanism of progressive RNA degradation by the archaeal exosome, *The Journal of*

biological chemistry 283, 14120-14131.
34. Walter, P., Klein, F., Lorentzen, E., Ilchmann, A., Klug, G., and Evguenieva-Hackenberg, E. (2006) Characterization of native and reconstituted exosome complexes from the hyperthermophilic archaeon Sulfolobus solfataricus, *Molecular microbiology 62*, 1076-1089.
35. The PyMOL Molecular Graphics System, Version 1.5.0.1 Schrödinger, LLC.
36. Chekanova, J. A., Shaw, R. J., Wills, M. A., and Belostotsky, D. A. (2000) Poly(A) tail-dependent exonuclease AtRrp41p from Arabidopsis thaliana rescues 5.8 S rRNA processing and mRNA decay defects of the yeast ski6 mutant and is found in an exosome-sized complex in plant and yeast cells, *The Journal of biological chemistry 275*, 33158-33166.
37. Burkard, K. T., and Butler, J. S. (2000) A nuclear 3'-5' exonuclease involved in mRNA degradation interacts with Poly(A) polymerase and the hnRNA protein Npl3p, *Molecular and cellular biology 20*, 604-616.
38. Lorentzen, E., Basquin, J., Tomecki, R., Dziembowski, A., and Conti, E. (2008) Structure of the active subunit of the yeast exosome core, Rrp44: diverse modes of substrate recruitment in the RNase II nuclease family, *Molecular cell 29*, 717-728.
39. Arcus, V. L., Backbro, K., Roos, A., Daniel, E. L., and Baker, E. N. (2004) Distant structural homology leads to the functional characterization of an archaeal PIN domain as an exonuclease, *The Journal of biological chemistry 279*, 16471-16478.
40. Levin, I., Schwarzenbacher, R., Page, R., Abdubek, P., Ambing, E., Biorac, T., Brinen, L. S., Campbell, J., Canaves, J. M., Chiu, H. J., Dai, X., Deacon, A. M., DiDonato, M., Elsliger, M. A., Floyd, R., Godzik, A., Grittini, C., Grzechnik, S. K., Hampton, E., Jaroszewski, L., Karlak, C., Klock, H. E., Koesema, E., Kovarik, J. S., Kreusch, A., Kuhn, P., Lesley, S. A., McMullan, D., McPhillips, T. M., Miller, M. D., Morse, A., Moy, K., Ouyang, J., Quijano, K., Reyes, R., Rezezadeh, F., Robb, A., Sims, E., Spraggon, G., Stevens, R. C., van den Bedem, H., Velasquez, J., Vincent, J., von Delft, F., Wang, X., West, B., Wolf, G., Xu, Q., Hodgson, K. O., Wooley, J., and Wilson, I. A. (2004) Crystal structure of a PIN (PilT N-terminus) domain (AF0591) from Archaeoglobus fulgidus at 1.90 A resolution, *Proteins 56*, 404-408.
41. Bunker, R. D., McKenzie, J. L., Baker, E. N., and Arcus, V. L. (2008) Crystal structure of PAE0151 from Pyrobaculum aerophilum, a PIN-domain (VapC) protein from a toxin-antitoxin operon, *Proteins 72*, 510-518.
42. Jeyakanthan, J., Inagaki, E., Kuroishi, C., and Tahirov, T. H. (2005) Structure of PIN-domain protein PH0500 from Pyrococcus horikoshii, *Acta crystallographica. Section F, Structural biology and crystallization communications 61*, 463-468.
43. Mattison, K., Wilbur, J. S., So, M., and Brennan, R. G. (2006) Structure of FitAB from Neisseria gonorrhoeae bound to DNA reveals a tetramer of toxin-antitoxin heterodimers containing pin domains and ribbon-helix-helix motifs, *The Journal of biological chemistry 281*, 37942-37951.
44. Glavan, F., Behm-Ansmant, I., Izaurralde, E., and Conti, E. (2006) Structures of the PIN domains of SMG6 and SMG5 reveal a nuclease within the mRNA surveillance complex, *The EMBO journal 25*, 5117-5125.
45. Takeshita, D., Zenno, S., Lee, W. C., Saigo, K., and Tanokura, M. (2007) Crystal structure of the PIN domain of human telomerase-associated protein EST1A, *Proteins 68*, 980-989.

46. Theobald, D. L., Mitton-Fry, R. M., and Wuttke, D. S. (2003) Nucleic acid recognition by OB-fold proteins, *Annual review of biophysics and biomolecular structure 32*, 115-133.
47. Frazao, C., McVey, C. E., Amblar, M., Barbas, A., Vonrhein, C., Arraiano, C. M., and Carrondo, M. A. (2006) Unravelling the dynamics of RNA degradation by ribonuclease II and its RNA-bound complex, *Nature 443*, 110-114.
48. Zuo, Y., Vincent, H. A., Zhang, J., Wang, Y., Deutscher, M. P., and Malhotra, A. (2006) Structural basis for processivity and single-strand specificity of RNase II, *Molecular cell 24*, 149-156.
49. Lebreton, A., Tomecki, R., Dziembowski, A., and Seraphin, B. (2008) Endonucleolytic RNA cleavage by a eukaryotic exosome, *Nature 456*, 993-996.
50. Schaeffer, D., Tsanova, B., Barbas, A., Reis, F. P., Dastidar, E. G., Sanchez-Rotunno, M., Arraiano, C. M., and van Hoof, A. (2009) The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities, *Nature structural & molecular biology 16*, 56-62.
51. Schneider, C., Leung, E., Brown, J., and Tollervey, D. (2009) The N-terminal PIN domain of the exosome subunit Rrp44 harbors endonuclease activity and tethers Rrp44 to the yeast core exosome, *Nucleic acids research 37*, 1127-1140.
52. Schaeffer, D., and van Hoof, A. (2011) Different nuclease requirements for exosome-mediated degradation of normal and nonstop mRNAs, *Proceedings of the National Academy of Sciences of the United States of America 108*, 2366-2371.
53. Wang, H. W., Wang, J., Ding, F., Callahan, K., Bratkowski, M. A., Butler, J. S., Nogales, E., and Ke, A. (2007) Architecture of the yeast Rrp44 exosome complex suggests routes of RNA recruitment for 3' end processing, *Proceedings of the National Academy of Sciences of the United States of America 104*, 16844-16849.
54. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis, *Journal of computational chemistry 25*, 1605-1612.
55. Malet, H., Topf, M., Clare, D. K., Ebert, J., Bonneau, F., Basquin, J., Drazkowska, K., Tomecki, R., Dziembowski, A., Conti, E., Saibil, H. R., and Lorentzen, E. (2010) RNA channelling by the eukaryotic exosome, *EMBO reports 11*, 936-942.
56. Mitchell, P., Petfalski, E., Houalla, R., Podtelejnikov, A., Mann, M., and Tollervey, D. (2003) Rrp47p is an exosome-associated protein required for the 3' processing of stable RNAs, *Molecular and cellular biology 23*, 6982-6992.
57. Stead, J. A., Costello, J. L., Livingstone, M. J., and Mitchell, P. (2007) The PMC2NT domain of the catalytic exosome subunit Rrp6p provides the interface for binding with its cofactor Rrp47p, a nucleic acid-binding protein, *Nucleic acids research 35*, 5556-5567.
58. Costello, J. L., Stead, J. A., Feigenbutz, M., Jones, R. M., and Mitchell, P. (2011) The C-terminal region of the exosome-associated protein Rrp47 is specifically required for box C/D small nucleolar RNA 3'-maturation, *The Journal of biological chemistry 286*, 4535-4543.
59. Midtgaard, S. F., Assenholt, J., Jonstrup, A. T., Van, L. B., Jensen, T. H., and Brodersen, D. E. (2006) Structure of the nuclear exosome component Rrp6p reveals an interplay between the active site and the HRDC domain, *Proceedings of the National Academy of Sciences of the United States of America 103*, 11898-11903.
60. Januszyk, K., Liu, Q., and Lima, C. D. (2011) Activities of human RRP6 and structure of

the human RRP6 catalytic domain, *RNA 17*, 1566-1577.
61. Zuo, Y., Wang, Y., and Malhotra, A. (2005) Crystal structure of Escherichia coli RNase D, an exoribonuclease involved in structured RNA processing, *Structure 13*, 973-984.
62. Zuo, Y., Zheng, H., Wang, Y., Chruszcz, M., Cymborowski, M., Skarina, T., Savchenko, A., Malhotra, A., and Minor, W. (2007) Crystal structure of RNase T, an exoribonuclease involved in tRNA maturation and end turnover, *Structure 15*, 417-428.
63. Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G., and Steitz, T. A. (1985) Structure of large fragment of Escherichia coli DNA polymerase I complexed with dTMP, *Nature 313*, 762-766.
64. Beese, L. S., and Steitz, T. A. (1991) Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism, *The EMBO journal 10*, 25-33.
65. Callahan, K. P., and Butler, J. S. (2008) Evidence for core exosome independent function of the nuclear exoribonuclease Rrp6p, *Nucleic acids research 36*, 6645-6655.
66. Lehner, B., and Sanderson, C. M. (2004) A protein interaction framework for human mRNA degradation, *Genome research 14*, 1315-1323.
67. Cristodero, M., Bottcher, B., Diepholz, M., Scheffzek, K., and Clayton, C. (2008) The Leishmania tarentolae exosome: purification and structural analysis by electron microscopy, *Molecular and biochemical parasitology 159*, 24-29.
68. Schilders, G., Raijmakers, R., Raats, J. M., and Pruijn, G. J. (2005) MPP6 is an exosome-associated RNA-binding protein involved in 5.8S rRNA maturation, *Nucleic acids research 33*, 6795-6804.
69. Milligan, L., Decourty, L., Saveanu, C., Rappsilber, J., Ceulemans, H., Jacquier, A., and Tollervey, D. (2008) A yeast exosome cofactor, Mpp6, functions in RNA surveillance and in the degradation of noncoding RNA transcripts, *Molecular and cellular biology 28*, 5446-5457.
70. Araki, Y., Takahashi, S., Kobayashi, T., Kajiho, H., Hoshino, S., and Katada, T. (2001) Ski7p G protein interacts with the exosome and the Ski complex for 3'-to-5' mRNA decay in yeast, *The EMBO journal 20*, 4684-4693.
71. Wang, L., Lewis, M. S., and Johnson, A. W. (2005) Domain interactions within the Ski2/3/8 complex and between the Ski complex and Ski7p, *RNA 11*, 1291-1302.
72. Andersen, G. R., Pedersen, L., Valente, L., Chatterjee, I., Kinzy, T. G., Kjeldgaard, M., and Nyborg, J. (2000) Structural basis for nucleotide exchange and competition with tRNA in the yeast elongation factor complex eEF1A:eEF1Balpha, *Molecular cell 6*, 1261-1266.
73. Andersen, G. R., Valente, L., Pedersen, L., Kinzy, T. G., and Nyborg, J. (2001) Crystal structures of nucleotide exchange intermediates in the eEF1A-eEF1Balpha complex, *Nature structural biology 8*, 531-534.
74. Pittman, Y. R., Valente, L., Jeppesen, M. G., Andersen, G. R., Patel, S., and Kinzy, T. G. (2006) Mg2+ and a key lysine modulate exchange activity of eukaryotic translation elongation factor 1B alpha, *The Journal of biological chemistry 281*, 19457-19468.
75. Kong, C., Ito, K., Walsh, M. A., Wada, M., Liu, Y., Kumar, S., Barford, D., Nakamura, Y., and Song, H. (2004) Crystal structure and functional analysis of the eukaryotic class II release factor eRF3 from S. pombe, *Molecular cell 14*, 233-245.
76. Atkinson, G. C., Baldauf, S. L., and Hauryliuk, V. (2008) Evolution of nonstop, no-go and nonsense-mediated mRNA decay and their termination factor-derived components,

BMC evolutionary biology 8, 290.
77. Kapp, L. D., and Lorsch, J. R. (2004) The molecular mechanics of eukaryotic translation, *Annual review of biochemistry 73*, 657-704.
78. Song, H., Mugnier, P., Das, A. K., Webb, H. M., Evans, D. R., Tuite, M. F., Hemmings, B. A., and Barford, D. (2000) The crystal structure of human eukaryotic release factor eRF1--mechanism of stop codon recognition and peptidyl-tRNA hydrolysis, *Cell 100*, 311-321.
79. Chekanova, J. A., Dutko, J. A., Mian, I. S., and Belostotsky, D. A. (2002) Arabidopsis thaliana exosome subunit AtRrp4p is a hydrolytic 3'-->5' exonuclease containing S1 and KH RNA-binding domains, *Nucleic acids research 30*, 695-700.
80. Yang, C. C., Wang, Y. T., Hsiao, Y. Y., Doudeva, L. G., Kuo, P. H., Chow, S. Y., and Yuan, H. S. (2010) Structural and biochemical characterization of CRN-5 and Rrp46: an exosome component participating in apoptotic DNA degradation, *RNA 16*, 1748-1759.
81. Chen, C. Y., Gherzi, R., Ong, S. E., Chan, E. L., Raijmakers, R., Pruijn, G. J., Stoecklin, G., Moroni, C., Mann, M., and Karin, M. (2001) AU binding proteins recruit the exosome to degrade ARE-containing mRNAs, *Cell 107*, 451-464.
82. Tomecki, R., Kristiansen, M. S., Lykke-Andersen, S., Chlebowski, A., Larsen, K. M., Szczesny, R. J., Drazkowska, K., Pastula, A., Andersen, J. S., Stepien, P. P., Dziembowski, A., and Jensen, T. H. (2010) The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L, *The EMBO journal 29*, 2342-2357.
83. Staals, R. H., Bronkhorst, A. W., Schilders, G., Slomovic, S., Schuster, G., Heck, A. J., Raijmakers, R., and Pruijn, G. J. (2010) Dis3-like 1: a novel exoribonuclease associated with the human exosome, *The EMBO journal 29*, 2358-2367.
84. Astuti, D., Morris, M. R., Cooper, W. N., Staals, R. H., Wake, N. C., Fews, G. A., Gill, H., Gentle, D., Shuib, S., Ricketts, C. J., Cole, T., van Essen, A. J., van Lingen, R. A., Neri, G., Opitz, J. M., Rump, P., Stolte-Dijkstra, I., Muller, F., Pruijn, G. J., Latif, F., and Maher, E. R. (2012) Germline mutations in DIS3L2 cause the Perlman syndrome of overgrowth and Wilms tumor susceptibility, *Nature genetics 44*, 277-284.
85. Schilders, G., van Dijk, E., and Pruijn, G. J. (2007) C1D and hMtr4p associate with the human exosome subunit PM/Scl-100 and are involved in pre-rRNA processing, *Nucleic acids research 35*, 2564-2572.
86. Lubas, M., Christensen, M. S., Kristiansen, M. S., Domanski, M., Falkenby, L. G., Lykke-Andersen, S., Andersen, J. S., Dziembowski, A., and Jensen, T. H. (2011) Interaction profiling identifies the human nuclear exosome targeting complex, *Molecular cell 43*, 624-637.
87. Wolfe, J. F., Adelstein, E., and Sharp, G. C. (1977) Antinuclear antibody with distinct specificity for polymyositis, *The Journal of clinical investigation 59*, 176-178.
88. Treadwell, E. L., Alspaugh, M. A., Wolfe, J. F., and Sharp, G. C. (1984) Clinical relevance of PM-1 antibody and physiochemical characterization of PM-1 antigen, *The Journal of rheumatology 11*, 658-662.
89. Reichlin, M., Maddison, P. J., Targoff, I., Bunch, T., Arnett, F., Sharp, G., Treadwell, E., and Tan, E. M. (1984) Antibodies to a nuclear/nucleolar antigen in patients with polymyositis overlap syndromes, *Journal of clinical immunology 4*, 40-44.
90. Reimer, G., Scheer, U., Peters, J. M., and Tan, E. M. (1986) Immunolocalization and partial characterization of a nucleolar autoantigen (PM-Scl) associated with

polymyositis/scleroderma overlap syndromes, *J Immunol 137*, 3802-3808.
91. Staals, R. H., and Pruijn, G. J. (2011) The human exosome and disease, *Advances in experimental medicine and biology 702*, 132-142.
92. Wu, C. J., Yang, X. F., McLaughlin, S., Neuberg, D., Canning, C., Stein, B., Alyea, E. P., Soiffer, R. J., Dranoff, G., and Ritz, J. (2000) Detection of a potent humoral response associated with immune-induced remission of chronic myelogenous leukemia, *The Journal of clinical investigation 106*, 705-714.
93. Faderl, S., Talpaz, M., Estrov, Z., and Kantarjian, H. M. (1999) Chronic myelogenous leukemia: biology and therapy, *Annals of internal medicine 131*, 207-219.
94. Fang, F., Hoskins, J., and Butler, J. S. (2004) 5-fluorouracil enhances exosome-dependent accumulation of polyadenylated rRNAs, *Molecular and cellular biology 24*, 10766-10776.
95. Kammler, S., Lykke-Andersen, S., and Jensen, T. H. (2008) The RNA exosome component hRrp6 is a target for 5-fluorouracil in human cells, *Molecular cancer research : MCR 6*, 990-995.

Chapter 2: Structural Studies of the Association of Rrp44 with the Exosome Core

Abstract

This chapter details efforts towards obtaining high resolution structural information on the association of Rrp44 with the eukaryotic exosome. It was determined that the N-terminal region of Rrp44 (amino acids 1-270) was necessary and sufficient for binding to exosome subunits Rrp41 and Rrp45. A ternary complex between Rrp41/Rrp45 and Rrp44 1-270 was then crystallized, but poor diffraction impeded structural determination. Other truncations of the N-terminal region of Rrp44 were attempted for crystallization in complex with Rrp41 and Rrp45, as well as Rrp44 homologues from other organisms, but crystals were not obtained. While our efforts were in progress, another group determined a structure of the full length Rrp44/Rrp41/Rrp45 ternary complex. Although this complex is helpful in elucidating how Rrp44 binds to two subunits of the exosome, it does not dock well into the EM reconstruction of the yeast Rrp44 bound exosome. Moreover, the structure indicated that the RNA binding channel of Rrp44 was blocked when bound to Rrp41/45, suggesting a mechanism of regulation of Rrp44 by the exosome. However, we solved the structure of the Apo Rrp44ΔPIN (amino acids 242-1001) and showed that the RNA binding channel is blocked in this structure as well, indicating that RNA binding causes a conformational change to open up the site, and that exosome binding is not involved. Overall, it appears that RNA could still potentially directly access Rrp44 without prior channeling through the exosome core, and that further studies are needed to determine the exact mode of interaction between Rrp44 and the complete exosome core during RNA degradation.

2.1. Introduction

In order to thoroughly understand the mechanism of RNA recruitment and processing by the eukaryotic exosome, it is critical to elucidate how Rrp44 binds to the exosome core. Although

the structure of the human core exosome has been solved (*1*), the human Rrp44 bound exosome would be a difficult structural target to pursue. At the time of the beginning of our experiments, the human Rrp44 was not known to associate with the human exosome core (*2*) yet later reports indicated that an association did exist (*3, 4*). The association of the human core exosome with Rrp44 is a more complex matter since three main version of Rrp44 are present in human cells (hDis3, hDis3L, and hDis3L2), as well as several potential isoform variants created by alternative splicing events (*3*). The human exosome has been shown to associate with a least two of the three known Rrp44 like proteins in human, hDis3 and hDis3L, (*3, 4*) yet the association appears to be relative weak, to the extent that it was hidden by previous experiments that used more stringent wash conditions (*2*). Due to the inherent difficulties of a human Rrp44 bound exosome, our studies focused mostly on the yeast exosome.

The yeast exosome associates tightly to Rrp44 (*5*), and the stable complex can be isolated using the TAP procedure, as demonstrated by the Ke group (*6*) as well as others (*7*), making it a better target to study the Rrp44-exosome association than the human exosome. Despite the fact that Rrp44 forms a stable complex with Rrp44, no crystal structures of the Rrp44 bound exosome have been forthcoming. The EM reconstruction of the yeast Rrp44 bound exosome by Hongwei Wang and the Ke group provided the first glimpse of the association of Rrp44 to the exosome core, and indicated a strong interaction of the N-terminal PIN domain of Rrp44 to the core subunits Rrp41 and Rrp45 (*6*). In order to more clearly understand this interaction, a higher resolution crystal structure of the Rrp44/Rrp45/Rrp41 sub-complex was needed. The EM reconstruction of the yeast Rrp44 bound exosome, as well as sequence conservation alignments were used to predict that a minimal region of Rrp44, amino acids 1-270 (hereby referred to as Rrp44ΔC), should be sufficient to bind to subunits Rrp41 and Rrp45, which was then confirmed by biochemical experiments. Our early efforts were focused on crystallizing the Rrp44ΔC/Rrp41/Rrp45 sub-complex.

2.2. Crystallization of an Rrp41/Rrp45/Rrp44 Head Ternary Complex

2.2.1. Methods

Plasmid Preparation

The coding regions of *S. cerevisiae* genes were PCR amplified from genomic DNA using Iproof high fidelity DNA polymerase enzyme (BioRad). A polycistronic plasmid containing the genes *Rrp41, Rrp45, and Rrp46* was created using overlap extension PCR (*8*), ligated into a pACYC184 vector containing a hexahistidine tag with a Tobacco Etch Virus Cleavage site, and tetracycline selection marker. The Rrp44 C-terminal deletion (ΔC, amino acids 1-270) was cloned using ligation independent cloning (LIC) (*9*) into a pLIC vector containing an N-terminal hexahistidine tag, a Tobacco Etch Virus Cleavage site, and a kanamycin selection marker. Plasmids were transformed into DH5α competent cells and grown on the kanamycin or tetracycline selective plates. DNA from single colonies was purified and all plasmids were sequenced to confirm the presence of the correct insert.

Protein Expression and Purification

For expression, plasmids were cloned into *E. coli* BL21* cells. Volumes of 3 – 4.5 L of antibiotic supplemented LB were inoculated with culture, grown at 37°C and 200 rpm to OD_{600} of 1.5 - 1.75, and induced with 1 mM IPTG at 18°C and 200 rpm overnight. Cell pellets were resuspended in Buffer A (20 mM Tris pH 7.5, 300 mM NaCl, 5 mM Imidazole, 5 mM betamercaptoethanol, 10% glycerol, and 1 mM PMSF) and were lysed by sonication. After centrifugation to pellet cell debris, supernatant was applied to Ni-NTA resin. The column was then washed with 10 column volumes of Buffer A containing 10 mM imidazole, and eluted with Buffer A supplemented with 300 mM Imidazole.

Complex Formation and Crystallization

After Ni-NTA purification, the Rrp44ΔC and the Rrp41/Rrp45 complex (Rrp46 is untagged and thus is absence from the complex) were then purified separately on a MonoQ column (GE Healthcare) in 20 mM Tris pH 7.5, 5 mM betamercaptoethanol, and 10% glycerol, and eluted using an increasing gradient of NaCl. The Rrp41/45 sub-complex and Rrp44 1-270 were ~90% pure as judged by SDS-PAGE analysis and Coomassie Blue staining (Fig. 2-1A-B). Protein concentrations were quantified by Bradford assay (*10, 11*).

For complex formation, Rrp41/45 and Rrp44ΔC were incubated in a 1:1.5 molar ratio in 20 mM Tris pH 7.5, 5 mM betamercaptoethanol, and 10% glycerol and purified on a MonoQ column, and peak fractions were analyzed. The ternary complex contained stoichiometric amounts of Rrp41, Rrp45, and Rrp44ΔC as judged by SDS-PAGE with Coomasie Blue staining, and was at least 95% pure (Fig.2-1C). Fractions containing the complex were pooled and concentrated to ~10 mg/mL for crystallization. Manual screening was completed using Hampton Screen I and II kits by mixing 1 µL of protein with 1 µL of well solution, and incubating at room temperature over a reservoir solution of 500 µL in large well hanging drop vapor diffusion trays. Initially, crystal showers appeared in 100 mM HEPES sodium, 20% PEG 4000, and 10% isopropanol (Fig.2-2A). After optimization and streak seeding, small, needle crystals of dimensions of about 5x5x50 μm^3 were produced in the final conditions of 100 mM HEPES sodium, 16.5% PEG 4000, and 10% isopropanol (Fig.2-2B).

2.2.2. Results

Crystals were tested for diffraction ability at the Advanced Photon Source at Argonne National Laboratory. In general, the crystals diffracted rather poorly, and at best a few diffraction spots were observed at ~ 5.5 Å (Fig. 2-2C). Thus, the diffraction was not of sufficient

Figure 2-1. SDS-PAGE analysis of Rrp41/45 and Rrp44 1-270. Proteins were purified by Ni-NTA chromatography followed by a MonoQ column, and peak fractions were analyzed on 10% polyacrylamide gels. A) The Rrp41 and Rrp45 sub-complex. B) Rrp44 amino acids 1-270, containing minor impurities. C) After incubating the Rrp41/45 sub-complex with Rrp44 1-270 and separating the mixture through a MonoQ column, the pure ternary complex elutes out in a single peak, followed by excess Rrp44 (not shown). BL is the before column loading sample, and M is the molecular weight marker.

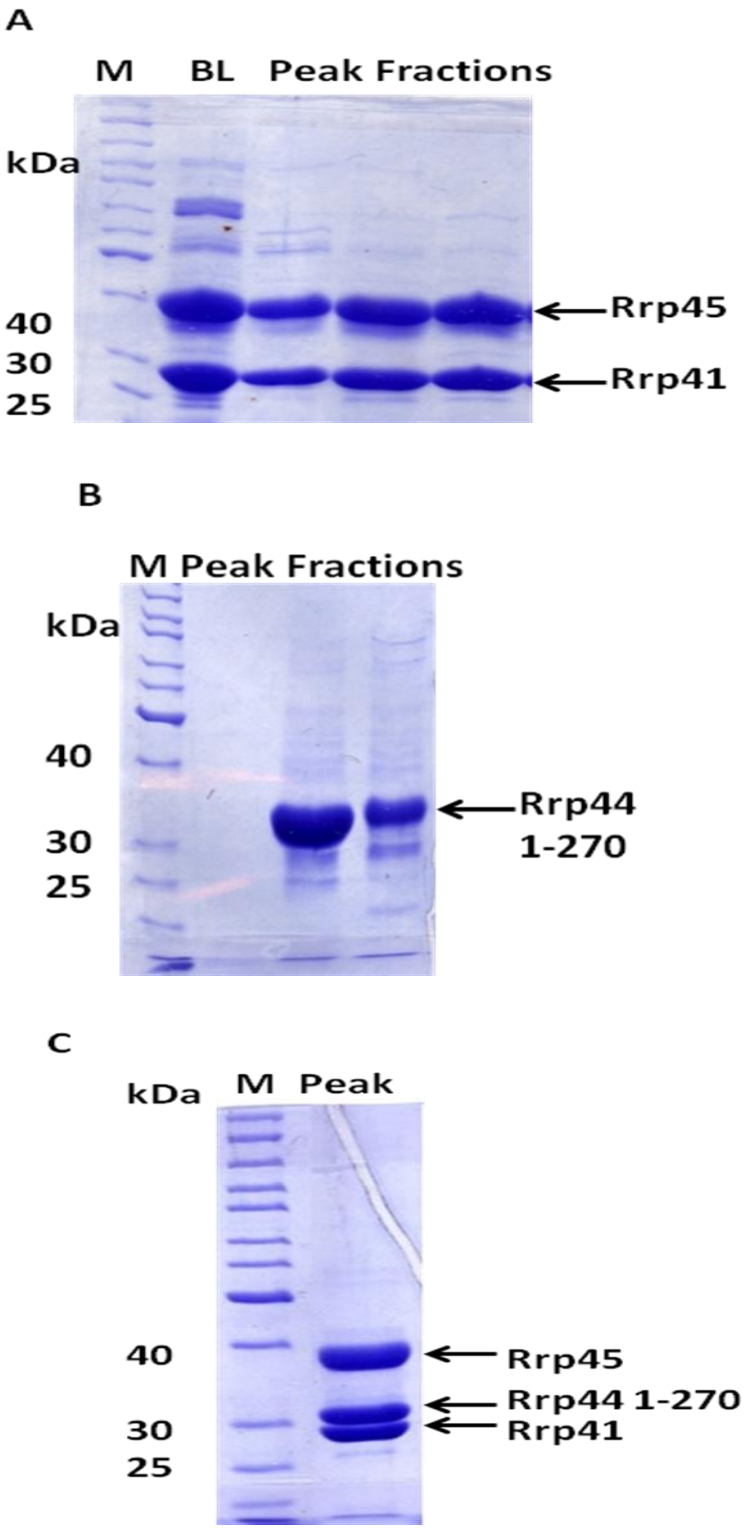
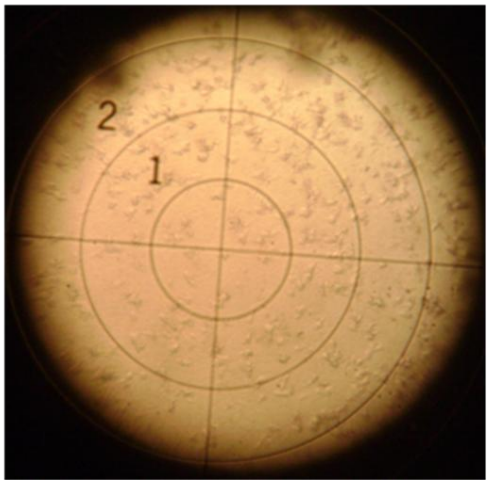
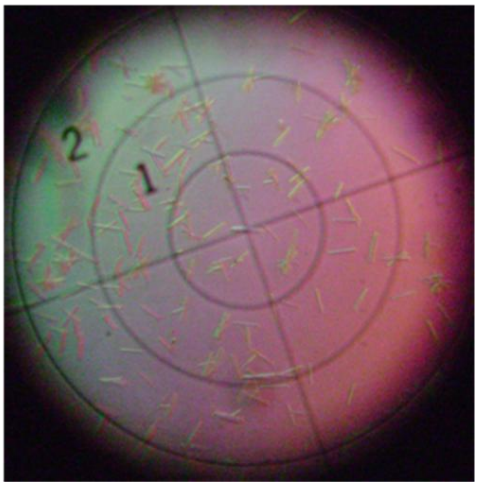

A
M
BL
Peak Fractions
kDa
40
30
25
Rrp45
Rrp41
B
M Peak Fractions
kDa
40
30
25
Rrp44
1-270
C
kDa
M
Peak
40
30
25
Rrp45
Rrp44 1-270
Rrp41

Figure 2.-2. Crystallization of the Rrp41/45/Rrp44ΔC ternary complex and diffraction analysis. A) Initial crystal showers were produced at a protein concentration of 10 mg/mL and mixing protein in a 1:1 ratio with crystallization solution containing 100 mM HEPES sodium pH 7.5, 16 – 20% PEG 4000, and 8-10% isopropanol in a hanging drop crystallization tray. B) Streak seeding the crystal showers into optimized conditions of 16.5% PEG 4000 and 10% isopropanol produced small needle crystals. The image shows crystals viewed under polarized light. C) Diffraction analysis of optimized crystals taken at the Advanced Photon Source at Argonne National Laboratory. Diffraction was insufficient for structural determination.

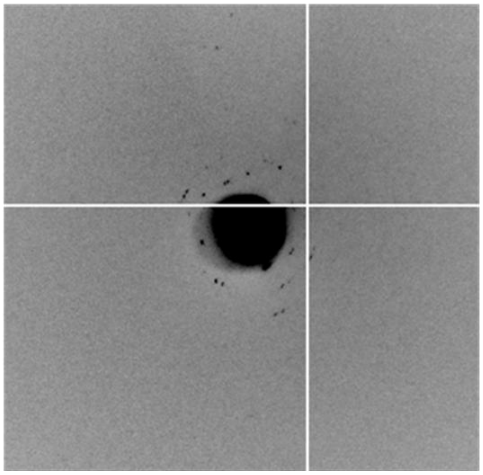
A



B



C



quality to collect a data set for structural determination.

2.3. Attempts at obtaining better diffracting crystals by Rrp44 engineering

Since the diffraction of the Rrp44/Rrp41/Rrp45 sub-complex crystals was insufficient for structural analysis, different truncations of Rrp44 were attempted in an effort to obtain better diffracting crystals.

2.3.1. Methods

Target Identification

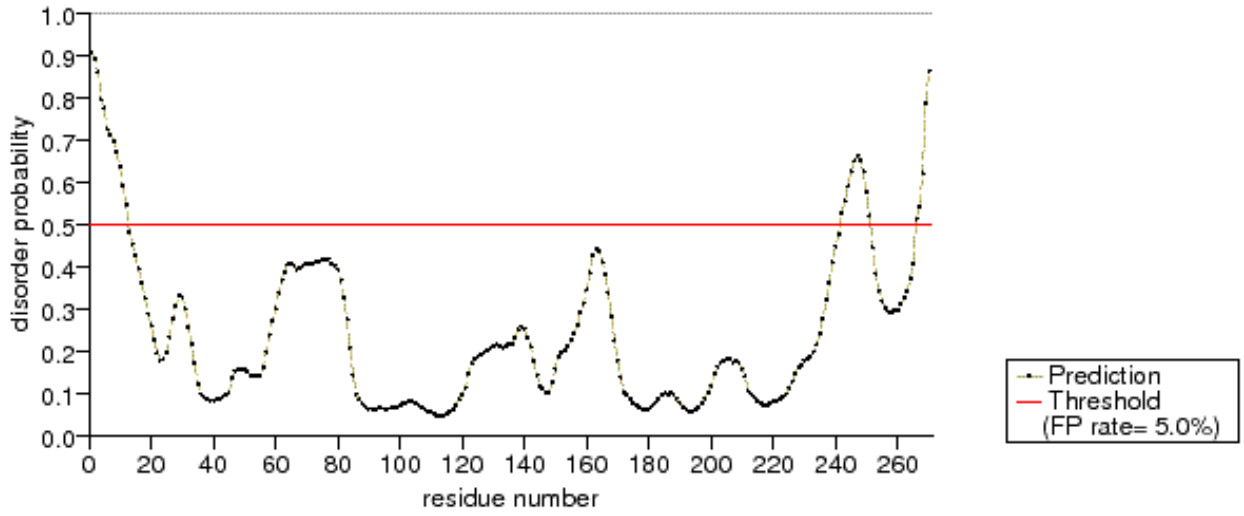
Several truncations (Table 2-1) were identified by various methods. The initial Rrp44ΔC (amino acids 1-270) had been identified by the EM reconstruction, and the remaining body region (270-1001) was also attempted. Additionally, a shorter truncation (1-242) was constructed based on the crystal structure of the C-terminal region of Rrp44 (*12*), and an even shorter version (17-242) removed the first several N-terminal amino acids that were predicted to be disordered using the PrDOS server (*13*) (Fig. 2-3A). A region predicted to comprise the PIN domain (86-198) was also attempted, but proved to be insoluble and therefore it was omitted from further analysis. Limited proteolysis of the complex (Fig. 2-3B) or Rrp44ΔC (not shown) with trypsin, chymotrypsin, and elastase and subsequent mass spectrometry analysis determined that the first 80 N-terminal amino acids were liable, as were amino acids 264-270 (Fig. 2-3B). Therefore, three constructs based on the limited proteolysis results (50-263, 80-263, and 1-263) were made. Finally, a minimal N-terminal region of Rrp44 (17-242) was fused to a maltose binding protein (MBP) moiety in an attempt to improve crystal packing (construct MBP 17-242).

Table 2-1. Rrp44 truncations attempted for use in complex formation with Rrp41/45.

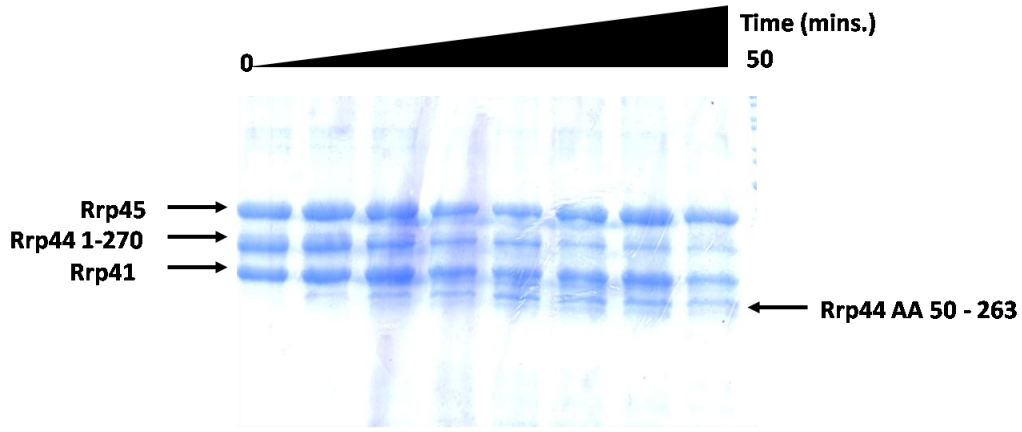
Rrp44 Truncation (Residue #)	Identity of Truncation	Complex Formation With Rrp41/45
1-270 (ΔC)	N-terminal head extended into CSD1	Yes
270-1001	C-terminal body	No
86-198	PIN Domain	No (insoluble)
50-263	Identified by limited proteolysis	No
80-263	Proteolyzed tail with smaller head	No
1-263	Full head and proteolyzed tail	Yes
1-242	Head without extension into CSD1	Yes
17-242	Removes predicted unstructured region	Yes
MBP 17-242	MBP fusion with unstructured region removed	No

Figure 2-3. Disordered prediction of Rrp44 and limited proteolysis of the Rrp41/45/Rrp44ΔC complex. A) The amino acid sequence of Rrp44 was input to the PrDOS server (*13*) to determine possible disordered regions that could be removed to aid in crystallization. Amino acids 1-17 at the extreme N-terminus as well a stretch between residues 240 and 270 had a high probability of being disordered, and insight from this prediction was used to make the Rrp44 17-242 construct. B) Limited proteolysis time course of the Rrp41/Rrp45/Rrp44 1-270 complex using a molar ratio of 1:1000 of trypsin to complex.

A


1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
disorder probability
0
20
40
60
80
100
120
140
160
180
200
220
240
260
residue number
Prediction
Threshold
(FP rate= 5.0%)

B


Time (mins.)
0
50
Rrp45
Rrp44 1-270
Rrp41
Rrp44 AA 50 - 263

Cloning, expression, purification, and complex formation analysis

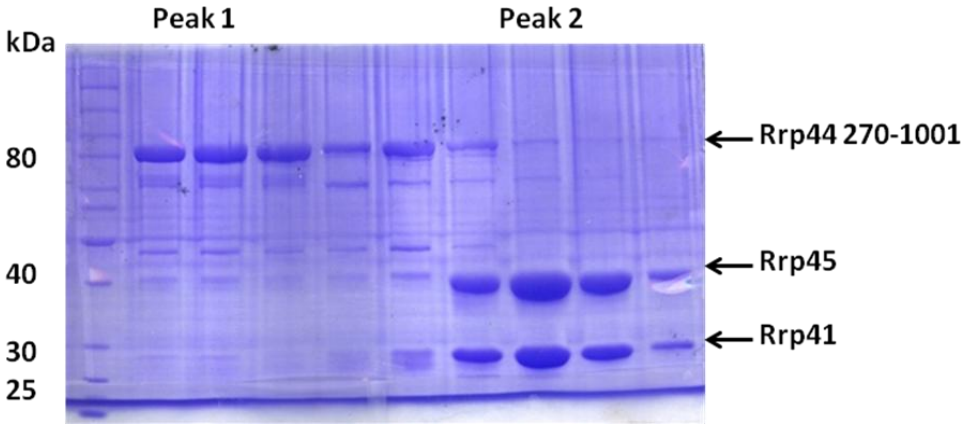
The truncations of Rrp44 (Table 2-1) were cloned into pLIC vectors and expressed in a manner similar to that of the Rrp44ΔC. The maltose binding protein (MBP) chimera construct was made by cloning Rrp44 17-242 into a modified pMAL-c2 vector, which contained a rigid penta-alanine linker between the fusion site and the protein of interest (*14*). The LB for the MBP-Rrp44 17-242 was supplemented with 2 g glucose during expression. Rrp44 plasmids were co-transformed with Rrp41/45 plasmid into BL21* cells, and selected for on plates containing kanamycin (or ampicillin for MBP Rrp44 17-242) and tetracycline. Large scale expression and purification was completed using Ni-NTA chromatography, and was performed in a manner similar to that of Rrp44ΔC (Section 2.2.1), with the exception of MBP Rrp44 17-242. Cell pellets of MBP Rrp44 17-242 were resuspended in buffer containing 20 mM Tris pH 7.5, 150 mM NaCl, 5 mM betamercaptoethanol, 10% glycerol, and 1 mM PMSF. The cell suspension was sonicated and the supernatant was bound to amylose resin, washed with 10 column volumes of buffer, and eluted with 10 mM maltose. Single step protein purifications were then run through a MonoQ column and eluted with a gradient of NaCl to determine if they formed a stable complex with Rrp41/45.

2.3.2. Results

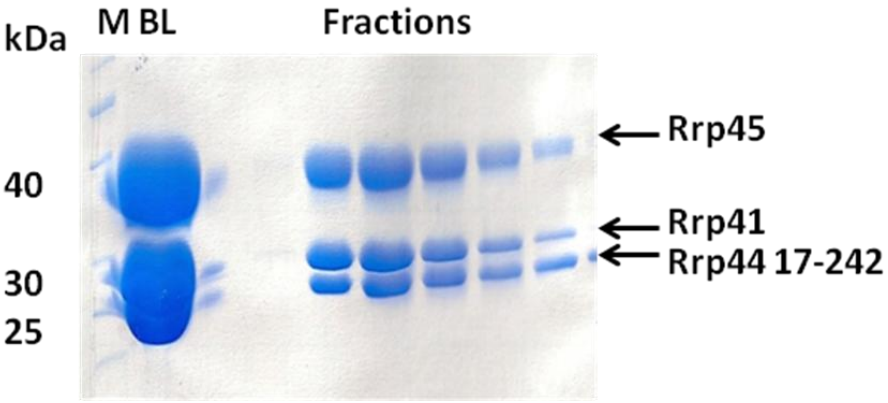
While all of the Rrp44 C-terminal truncations (amino acids 1-242 and 1-263) and the minimal N-terminal truncation amino acids 17-242 were found to form stable complexes with Rrp41 and Rrp45, more extreme N-terminal truncations (aa 50-263, 80-263, and 270-1001) did not form stable complexes (see Fig. 2-4A for the 270-1001 result), suggesting that main source of

Figure 2-4. Association of truncated Rrp44 with Rrp41/45. A) The C-terminal region of Rrp44 (amino acids 270-1001) was mixed with Rrp41/45 complex, and separated on a MonoQ column. The chromatography resulted in two main peaks (not shown). The first peak contained only Rrp44 270-1001, while the second peak contained only Rrp41/45. B) and C) Association of Rrp44 C-terminal truncations with Rrp41/45. Complexes were purified by Ni-NTA, MonoQ, and gel filtration chromatography and analyzed on 10% polyacrylamide gels. B) Rrp44 17-242 + Rrp41/45. C) Rrp44 1-242 + Rrp41/45. Rrp41 and Rrp44 1-242 run in overlapping positions. Abbreviations: M: molecular weight marker. BL: protein sample before loading on the column.

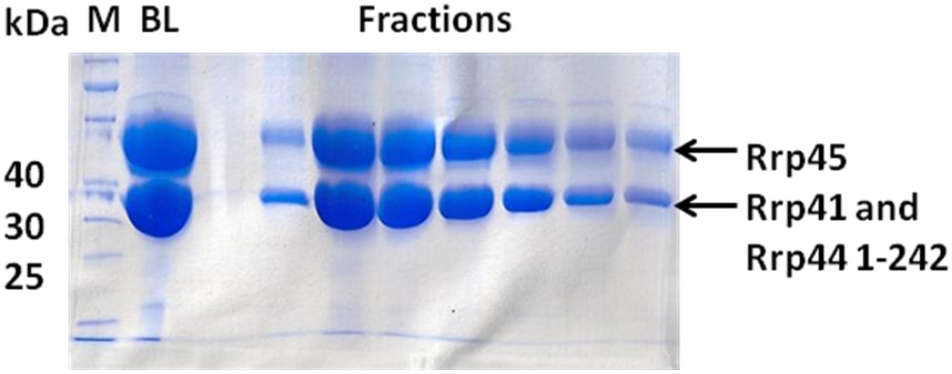
A


Peak 1
Peak 2
kDa
80
40
30
25
Rrp44 270-1001
Rrp45
Rrp41

B


kDa
M BL
Fractions
40
30
25
Rrp45
Rrp41
Rrp44 17-242

C


kDa
M
BL
Fractions
40
30
25
Rrp45
Rrp41 and
Rrp44 1-242

association of Rrp44 with the exosome resided in the PIN domain of Rrp44, as was predicted from the EM reconstruction. In addition, the MBP Rrp44 17-242 construct was found to dissociate from Rrp41 and Rrp45 during the initial amylose column purification; presumably this occurred because the MBP moiety was fused to the N-terminus of Rrp44 and thus sterically inhibited its interaction with Rrp41 and Rrp45.

2.4. Crystallization screening of truncated Rrp44 complexes

All of the Rrp44 truncations that successfully formed a complex with Rrp41 and Rrp45 (aa 1-242, 1-263, and 17-242) were considered potential candidates for crystallization. Large scale expression and multi-step purification (Ni-NTA, MonoQ, and size exclusion chromatography) was completed for these complexes (Fig. 2-4B-C for 1-242 and 17-242 complexes). They were then concentrated to 10 – 20 mg/mL and nano-crystal screening using sitting drop trays and volumes of ~0.1-0.3 µL using the Phoenix Crystallization Robot at the Weill Center for Molecular and Cell Biology. However, no obvious crystal hits were obtained from screening.

Although MBP Rrp44 17-242 did not form a complex with Rrp41/45, it still served as a potential crystallization target because the structure of the N-terminal PIN domain of Rrp44 was unknown at that time. MBP Rrp44 17-242 was further purified on a MonoQ column (Fig. 2-5A), which removed excess MBP that inevitably results during over-expression. The protein was concentrated for crystallization screening in buffer containing 20 mM Tris pH 8.0, 150 mM NaCl, 3 mM $MnCl_2$, and 1 mM DTT. Initial screening at protein concentrations of ~10 mg/mL produced only clear drops, most likely because the MBP tag is a solubility enhancer and thus prevents the protein from precipitating out of solution. The remainder of the protein stock was

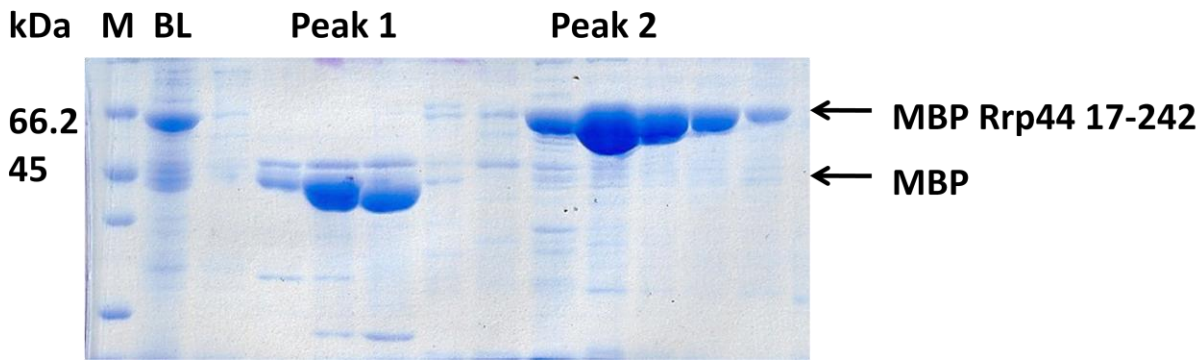
therefore further concentrated to 28 mg/mL and tested for crystallization using nano scale sitting drop trays. After incubation of approximately 1 month, crystals appeared in a precipitation condition containing 0.2 M NaCl, 0.1 M HEPES pH 7.5, and 10% v/v 2-propanol (Fig. 2-5B). The entirety of these crystals were washed and run on SDS-PAGE, and appeared to contain the protein of interest (Fig. 2-5C). However, crystals were not readily reproducible using large scale crystallization optimization with different batches of protein, even at higher protein concentrations up to 46 mg/mL. It is possible that crystallization was simply not reproducible when scaling up to large scale screens. Another possibility is that differences in purity between the two protein batches caused crystallization problems. It is also possible that free MBP impurities resulted in prior crystallization, and subsequent batches were purified to a greater degree to remove free MBP.

2.5. Hetero-assembly of Rrp44 from different organisms with yeast Rrp41/Rrp45

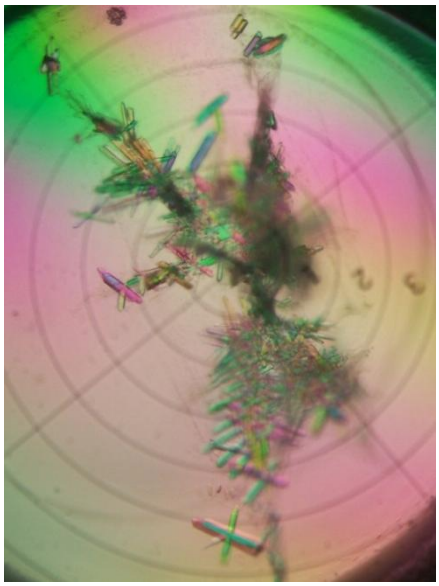
In a combined effort to search for new crystallization targets of the Rrp44/Rrp41/Rrp45 complex as well as to probe the evolutionary conservation of interaction of Rrp44 with Rrp41 and Rrp45, purification of Rrp44 proteins from organisms other than *S. cerevisiae* were attempted and tested for their ability to form a complex with yeast Rrp41/45. The organisms that served as a source of Rrp44 proteins were *Schizosacchromyces pombe* and *Homo sapiens*, and regions homologous to the crystallized fragment of yeast Rrp44 (aa 1-270) were determined by sequence alignment (Fig. 2-6). At the time of this research, human Rrp44 (hDis3) had been shown to complement yeast cells depleted of Rrp44 (*15*), but it had not yet been demonstrated that it bound to the human exosome core.

Figure 2-5. Purification and Crystallization of MBP Rrp44 17-242. A) MonoQ purification. Pure MBP Rrp44 17-242 was contained in a late eluting peak, while free MBP bound less readily to the column and eluted earlier. B) Crystals produced in 0.2 M NaCl, 0.1 M HEPES pH 7.5, and 10% v/v 2-propanol at protein concentration of 28 mg/mL. C) Washing the crystals and running them on a gel produced a band indicative of the molecular weight of the fusion protein. M is the molecular weight marker, SN is the supernatant produced after centrifuging crystals, W1 and W2 are washes 1 and 2, and Xtal is the crystal.

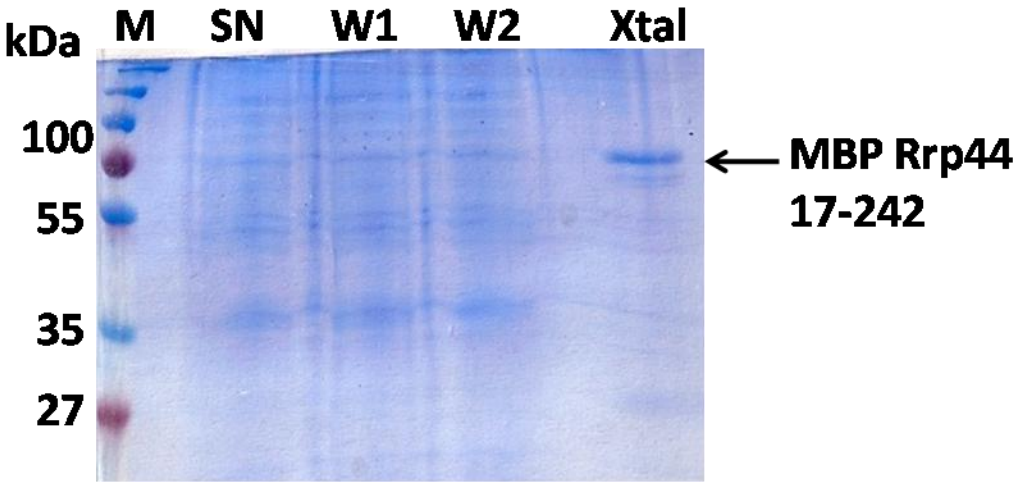
A


kDa
M
BL
Peak 1
Peak 2
66.2
45
MBP Rrp44 17-242
MBP

B



C


kDa
M
SN
W1
W2
Xtal
100
55
35
27
MBP Rrp44
17-242

2.5.1. Methods

Cloning

Human Rrp44 amino acids 1-215 was PCR amplified from cDNA from both cardiomyocytes and MCF-7 cells (kindly provided by the Kraus lab) using primers with a 5' end BamHI site and 3' end XhoI site. While initial amplification attempts using a standard PCR program were unsuccessful, a touchdown PCR program (*16*) was used to successfully amplify the region. The PCR product was digested with XhoI and BamHI and ligated into a vector containing an N-terminal hexahistidine SUMO fusion and kanamycin selective marker (pSUMO, kindly provided by the lab of Holger Sondermann). Ligated products were transformed into DH5α cells and grown on kanamycin selective plates. Single colonies were picked and mini-preped to isolate plasmids, and plasmids were confirmed to contain the insert of interest by DNA sequencing.

Interestingly, sequencing revealed cell specific isoforms of human Rrp44. Plasmids constructed using PCR products from reactions containing cDNA from cardiomyocytes as a template contained the full length N-terminal region of human Rrp44 (also known as hDis3, and listed in pubmed as isoform a), while those that used MCF-7 cDNA as a template contained a shorter N-terminal region (Fig. 2-7A) that also had about 20 different amino acids (listed in pubmed as isoform b), which appears to result from alternative splicing, and has no relation to the other "canonical" human isoforms (hDis3L1 or hDis3L2). This region is well conserved in higher eukaryotes (Fig. 2-7B), but its function is unknown. Notably, in isoform b there is a deletion for a residue involved in endonuclease activity in yeast Rrp44 (D91 using yeast

numbering). Plasmids encoding the full length human Rrp44 and the N-terminal truncation amino acids 215-end were also constructed in a similar fashion. Plasmids were transformed into BL21* cells for use in protein purification. *S. pombe* Rrp44 amino acids 1-235 was PCR amplified from genomic DNA (kindly provided by the laboratory of Jeff Pleiss), using primers containing a 5' end BamHI site and 3' end XhoI site. A bridging PCR method with overlapping primers was used to remove introgenic regions from the PCR product, and touchdown PCR was successfully used to amplify the protein coding region. *S. pombe* Rrp44 1-235 was ligated into the pSUMO vector and confirmed by sequencing in a manner similar as that described for human Rrp44, before being transformed into BL21* cells in preparation for protein expression.

Protein Expression and Purification

Expression and purification steps were similar for all proteins. Cultures were grown in 3 L of LB media at 37 °C until reaching OD600 of 1.7, at which time they were induced with 1 mM IPTG overnight at 18 °C. Proteins were purified by Ni-NTA chromatography using procedures similar to those mentioned earlier. While purifications of *S. pombe* Rrp44 amino acids 1-235 and human Rrp44 1-215 isoform a were successful, isoform b appeared to express poorly, and human Rrp44 full length and C-terminal truncation both appeared to be proteolyzed into several bands during purification; therefore, these proteins were omitted from further analysis. The SUMO fusion tag was cleaved by overnight incubation with yeast Sm3t SUMO protease at 4° C. During cleavage, proteins were dialyzed into buffer containing 20 mM Tris pH 8.0, 50 mM NaCl, 5 mM betamercaptoethanol, and 10% glycerol in order to lower the salt concentration and to remove imidazole. Imidazole free proteins were then run through a second nickel column to remove the hexahistidine SUMO tag, loaded onto a heparin column, and eluted with a gradient of NaCl. Purified proteins were analyzed by SDS-PAGE (Fig. 2-8) and concentrated.

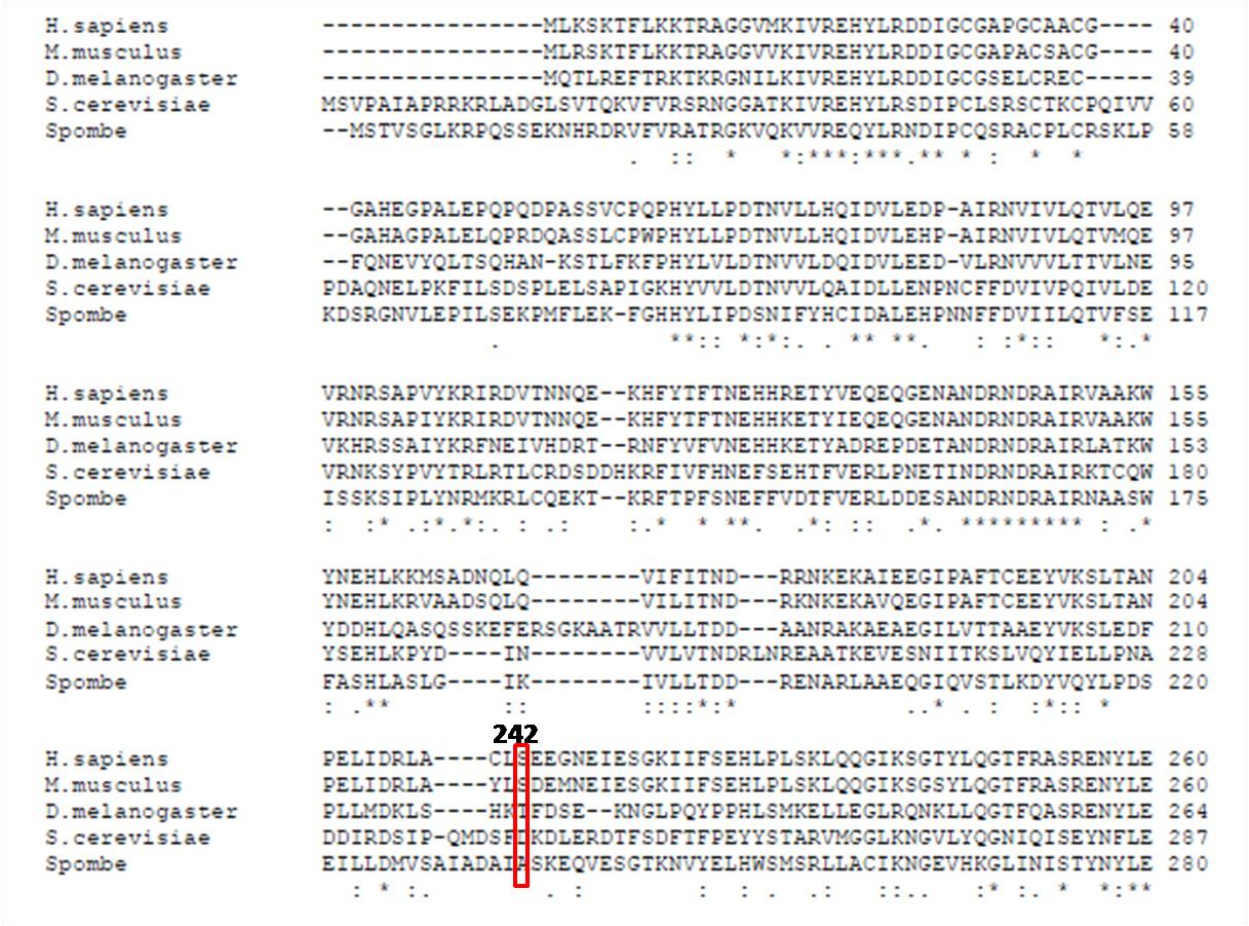


Figure 2-6. Sequence alignment of the N-terminal region of Rrp44 from various eukaryotes. Sequences from several common organisms were aligned using the ClustalW server (*17*). The border of the N-terminal region (amino acid 242 in *S. cerevisiae*) is boxed in red.

A

```
hRrp44isoformA      MLKSKTFLKKTRAGGVMKIVREHYLRDDIGCGAPGCAACGGAHEGPALEPQPQDPASSVC 60
hRrp44isoformB      MLKSKTFLKKTRAGGVMKIVREHYLRDDIGCGAPGCAACGGAHEGPALEPQPQDPASSVC 60
                    ************************************************************

hRrp44isoformA      PQPHYLLPDTNVLLHQIDVLEDPAIRNVIVLQTVLQEVRNRSAPVYKRIRDVTNNQEKHF 120
hRrp44isoformB      PQPHYLLPDTNVLLHQIVSAWRPGT-----WASVASSLRLPGS----------------- 98
                    *****************     *.        :* ..:*  .:

hRrp44isoformA      YTFTNEHHRETYVEQEQGENANDRNDRAIRVAAKWYNEHLKKMSADNQLQVIFITNDRRN 180
hRrp44isoformB      --------LETYVEQEQGENANDRNDRAIRVAAKWYNEHLKKMSADNQLQVIFITNDRRN 150
                            ************************************************

hRrp44isoformA      KEKAIEEGIPAFTCEEYVKSLTANPELIDRLACLSEEGNEIESGKIIFSEHLPLSKLQQG 240
hRrp44isoformB      KEKAIEEGIPAFTCEEYVKSLTANPELIDRLACLSEEGNEIESGKIIFSEHLPLSKLQQG 210
                    ************************************************************

hRrp44isoformA      IKSGTYLQGTFRASRENYLEATVWIHGDNEENKEIILQGLKHLNRAVHEDIVAVELLPKS 300
hRrp44isoformB      IKSGTYLQGTFRASRENYLEATVWIHGDNEENKEIILQGLKHLNRAVHEDIVAVELLPKS 270
                    ************************************************************

hRrp44isoformA      QWVAPSSVVLHDEGQNEEDVEKEEERERMLKTAVSEKMLKPTGRVVGIIKRNWRPYCGML 360
hRrp44isoformB      QWVAPSSVVLHDEGQNEEDVEKEEETERMLKTAVSEKMLKPTGRVVGIIKRNWRPYCGML 330
                    ************************* ************************************
```

B

```
H.sapiensIsoformA     PQPHYLLPDTNVLLHQIDVLEDPAIRNVIVLQTVLQEVRNRSAPVYKRIRDVTNNQEKHF 120
B.taurus              PQPHYLLPDTNVLLHQIDVLEDPAIRNVIVLQTVLQEVRNRSAPVYKRIRDMTNNQEKHF 120
M.musculus            PWPHYLLPDTNVLLHQIDVLEHPAIRNVIVLQTVMQEVRNRSAPIYKRIRDVTNNQEKHF 120
X.laevis              TPPHYLLPDTNVLLHQIDILEDPIIKNLIILQTVLQEVRSRSAPVHKRVKDLLHNPDKCF 117
                      . ***************:**.* *:*:*:****:****.****::**::*: :* :* *

H.sapiensIsoformA     YTFTNEHHRETYVEQEQGENANDRNDRAIRVAAKWYNEHLKKMSADNQLQVIFITNDRRN 180
B.taurus              YTFTNEHHRETYVEQLQGENSNDRNDRAIRVAAKWYNEHLKNMSAENRLQVIFITNDRKN 180
M.musculus            YTFTNEHHKETYIEQEQGENANDRNDRAIRVAAKWYNEHLKRVAADSQLQVILITNDRKN 180
X.laevis              YTFTNEHHRETYIEQEQGENANDRNDRAIRTATKWYNKHLKKSPSMANTQVILITNDRRN 177
                      ********:**:** ****:*********.*:****:***. .:  . ***:*****:*
```

Figure 2-7. Sequence conservation between two isoforms of human Rrp44 (hDis3). A) Alignment of hRrp44 isoforms. B) Alignment of Rrp44 in higher eukaryotes. The boxed region is well conserved in higher eukaryotes, but differs between the two human isoforms. Alignments were performed with ClustalW (*17*).

Complex Formation Studies

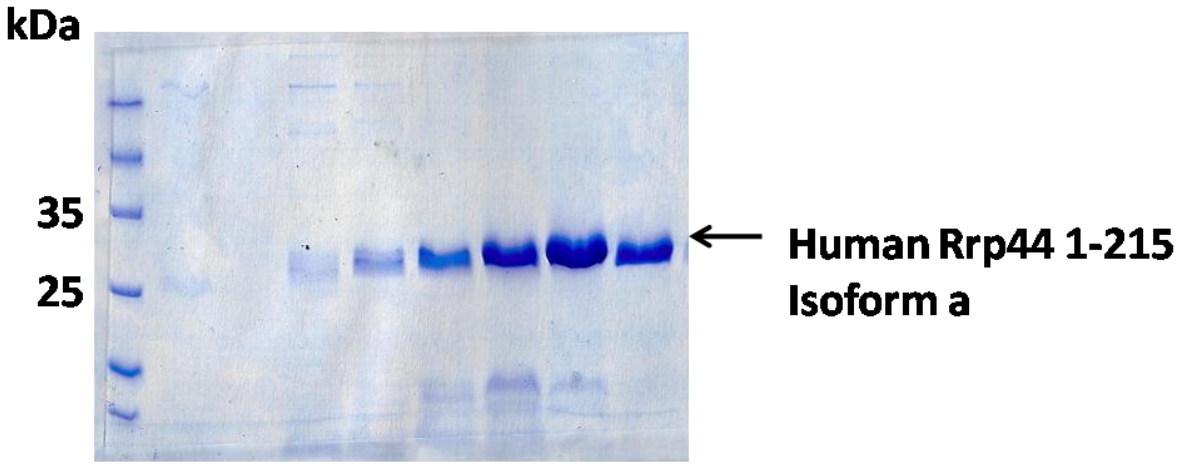
Proteins were used in complex formation assays with Rrp41/45. Human Rrp44 1-215 or *S. pombe* Rrp44 1-235 were separately mixed in a 1.5:1 molar ratio with *S. cerevisiae* Rrp41/45 and incubated for 30 minutes on ice. The mixtures were then loaded onto a Superdex S200 size exclusion column and run in buffer containing 20 mM Tris pH 8.0, 150 mM NaCl, and 1 mM DTT at a flow rate of 0.25 ml/min. Peak fractions were analyzed by SDS-PAGE. Human Rrp44 1-215, *S. pombe* Rrp44 1-235, and Rrp41/45 were also run alone using the same column for comparison, and chromatographs (not shown) for runs of the single proteins were used as elution volume references.

2.5.2. Results

Chromatographs of both assays consisted of two prominent peaks (Fig. 2-9). In both cases, the first peak ran in a position coincident with that of Rrp41/45 alone (not shown) and subsequent SDS-PAGE analysis indicated that the peak contained only Rrp41/45, while the second peak eluted at a volume representative of the Rrp44 protein and contained either only human Rrp44 1-215 or *S. pombe* Rrp44 1-235 (Fig. 2-9A-B). These results indicate that the association of Rrp44 with the exosome core is different enough to the extent that heterogeneous assembly failed. A recent report indicates that hDis3 can bind to the human exosome but the association is very salt sensitive (stable at only 75-100 mM of NaCl) (*4*), so it is possible that the salt concentration used in this study (150 mM NaCl) was too high. However, this seems unlikely, given that yeast Rrp44 does not dissociate from the yeast exosome at NaCl concentrations of 500 mM. It is more likely that the human exosome associates with Rrp44 in a manner different than the yeast exosome. While human Rrp44 (hDis3) can complement Rrp44 yeast depleted cells (*15*), it is possible that this does not require binding of Rrp44 to the exosome. The human exosome could

Figure 2-8. SDS-PAGE analysis of *S. pombe* Rrp44 Truncations. Gels represent protein purity after two step purification (Ni-NTA and heparin). A) *S. pombe* amino acids 1-235. B) Human isoform a amino acids 1-215

A


kDa
35
25
Human Rrp44 1-215
Isoform a

B

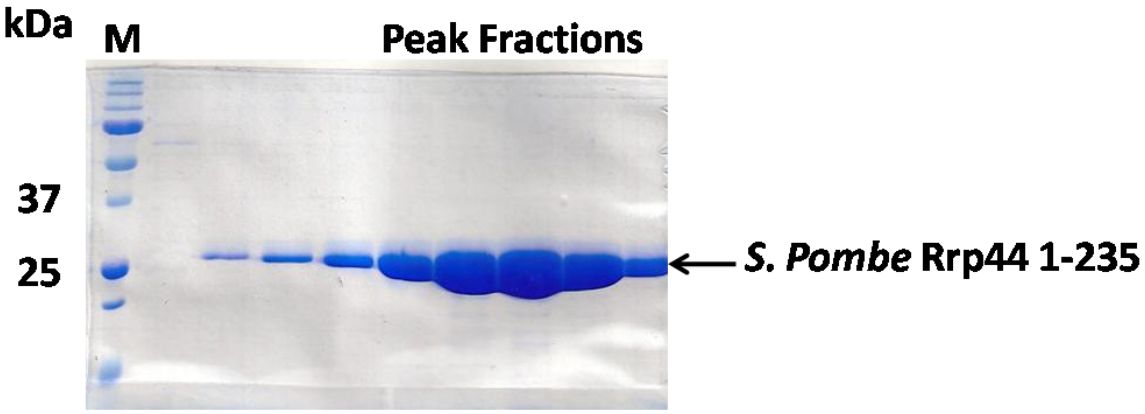
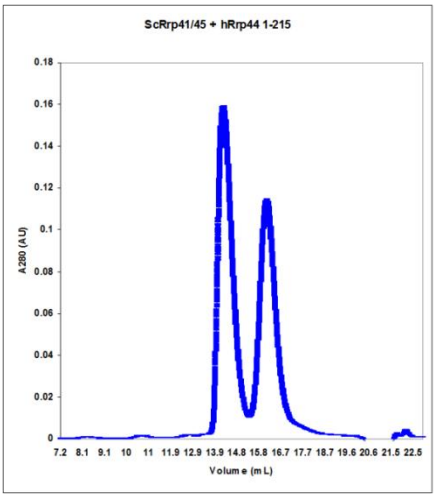
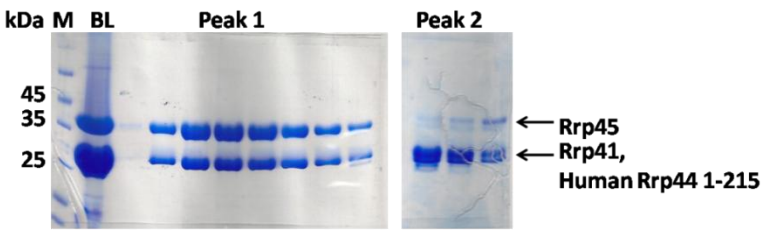

kDa
M
Peak Fractions
37
25
S. Pombe Rrp44 1-235

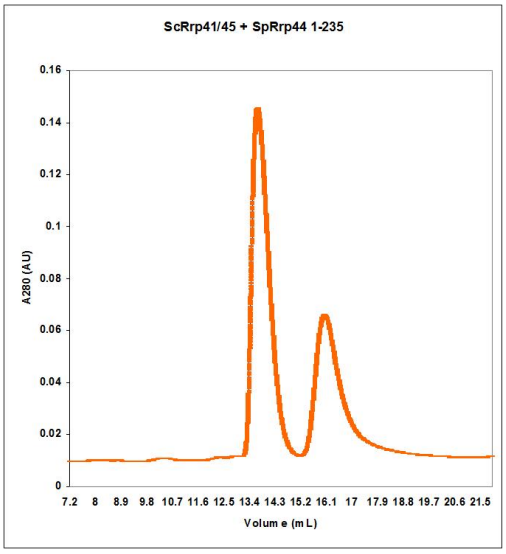
Figure 2-9. Rrp44 complex formation assayed by size exclusion chromatography. Rrp44 proteins were mixed in a 1.5:1 molar ratio with Rrp41/45 and run through a Superdex S200 size exclusion column. A) Human Rrp44 1-215 + Rrp41/45. B) *S. pombe* Rrp44 1-235 + Rrp41/45. In both cases, two peaks appeared in the chromatography. The first peak was centered at an elution volume of about 14 mL, and contained only Rrp41/45 as analyzed by SDS-PAGE. The second peak was centered around and elution volume of about 16 mL, the same elution volume for the Rrp44 proteins alone (not shown) and contained only the Rrp44 protein on SDS-PAGE. C) Size exclusion chromatography analysis of the *S. cerevisiae* Rrp44 1-270 + Rrp41/45 complex for means of comparison. In this chromatograph, the first peak contains the ternary complex, while the small second peak contains excess Rrp44.

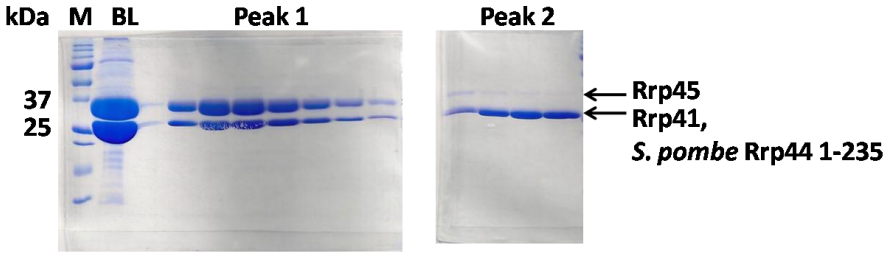
A


ScRrp41/45 + hRrp44 1-215
A280 (AU)
Volume (mL)

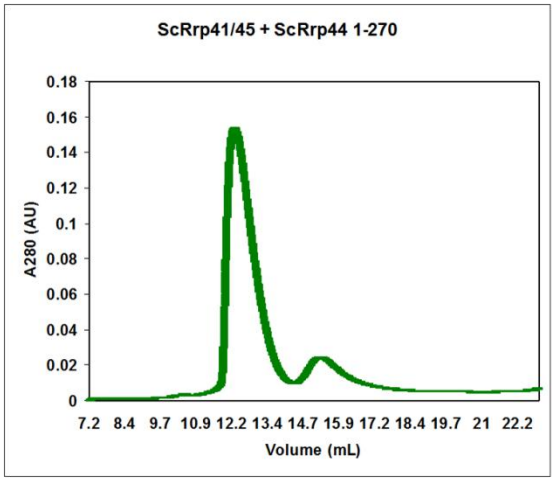

kDa M BL
Peak 1
Peak 2
45
35
25
Rrp45
Rrp41,
Human Rrp44 1-215

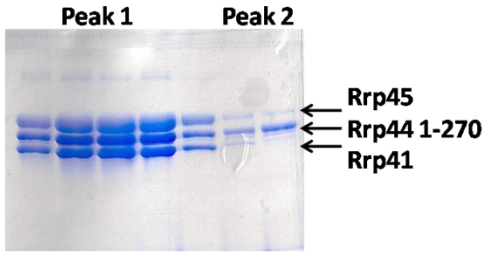
B


ScRrp41/45 + SpRrp44 1-235
A280 (AU)
Volume (mL)


kDa M BL
Peak 1
Peak 2
37
25
Rrp45
Rrp41,
S. pombe Rrp44 1-235

C


ScRrp41/45 + ScRrp44 1-270
A280 (AU)
Volume (mL)


Peak 1
Peak 2
Rrp45
Rrp44 1-270
Rrp41

also require a larger region of Rrp44 to bind to in order to form a stable association with the exosome core than just the N-terminal region of Rrp44, or make contacts with different exosome core subunits. Therefore, a continuation along this line of inquiry was planned to clone out Rrp41 and Rrp45 from human and *S. pombe*, purify them, and to test assembly of the organism specific ternary complexes with Rrp44 for use in crystallization screening.

2.6. Ideas on a complex containing the full length Rrp44 and Rrp41/45

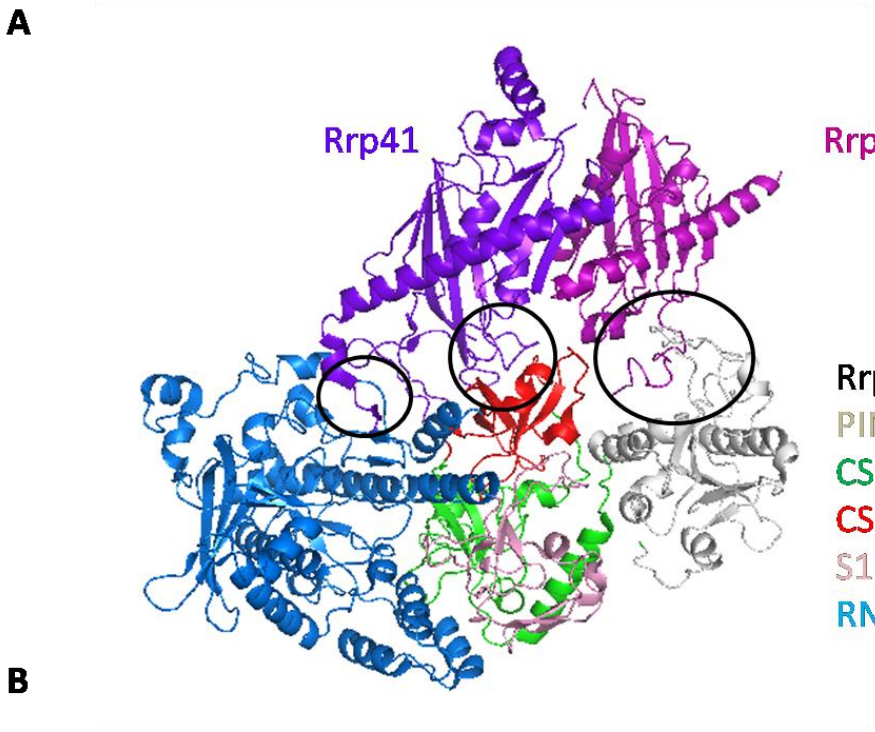
While work on cloning of the human and *S. pombe* Rrp41 and Rrp45 commenced, an attempt to form the *S. cerevisiae* ternary complex using the full length Rrp44 was planned. Initially, the minimal ternary complex containing only the N-terminal region of Rrp44 was rigorously pursued because Rrp44 was predicted from the EM reconstruction results to contain a flexible linker joining the N and C terminal regions (*6*), which could make crystallization of the full length Rrp44 difficult. On the other hand, since structural studies of the minimal ternary complex of the N-terminal region of Rrp44 bound to subunits Rrp41 and Rrp45 at best only produced poorly diffracting crystals (Section 2.2), it was envisioned that using a complex containing the full length version of Rrp44 bound to Rrp41/45 could provide more regions of protein contact, which could improve crystal packing and ultimately provide better diffracting crystals. Work was begun on purifying the yeast full length Rrp44 + Rrp41/45 ternary complex; however, before crystallization screening of the complex was completed, another group solved a similar structure and published their report (*18*). Unfortunately, this report rendered crystallization of both MBP Rrp44 17-242 and the Rrp44ΔC/Rrp41/Rrp45 ternary complex superfluous. Since cloning of human and *S. pombe* Rrp41 and Rrp45 subunits had not yet begun, it was abandoned.

2.7. Relation of the published ternary structure to our efforts

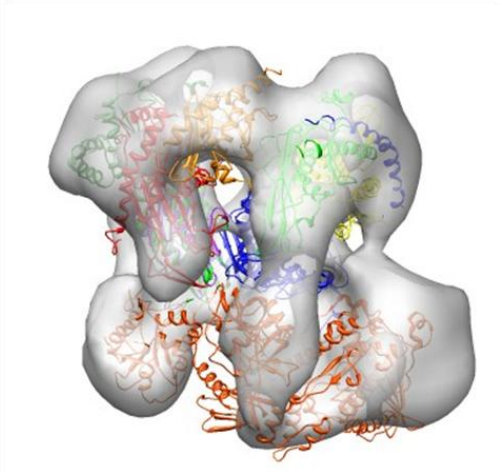
Since the Rrp44/Rrp41/Rrp45 ternary complex structure has been published by the Conti group (*18*), it will only be covered briefly and with an emphasis on its relation to prior work completed by the Ke group. Overall, experiments conducted by the Conti group concerning the association of Rrp44 with the exosome core agreed well with our results. Similar to our results, they found a stable association of Rrp44 with the exosome subunits Rrp41 and Rrp45, and they reported that deleting the PIN domain of Rrp44 resulted in a loss of association with the exosome core (*18*). For crystallographic studies, the Conti group used the *S. cerevisiae* Rrp44 truncation amino acids 25-1001, which was reportedly determined using limited proteolysis of the Rrp44 PIN domain, which was used within the physiological C-terminal region of the protein, and crystallized in complex with full length versions of *S. cerevisiae* Rrp41 and Rrp45 (*18*) .

The PIN domain of Rrp44 serves as the main site of interaction with the exosome core. PIN domain residues 37-71 interact with Rrp41 residues N-terminal residues (Fig. 2-10A), and the interaction is held together by the formation of salt bridges of residues Glu10, Arg18, and Glu21 of Rrp41 to residues Arg133, Asp44, and Arg42 of Rrp44, respectively (*18*). Although our results agree with those of the Conti group that N-terminus of Rrp44 was necessary and sufficient for association with the Rrp41/45 sub-complex, the crystal structure of the ternary complex revealed that the C-terminal region of Rrp44 was found to bind to regions on both Rrp41 and Rrp45 as well. The CSD2 domain of Rrp44 was found to interact with Rrp45 via electrostatic interactions between charged residues. The RNB domain of Rrp44 binds to two surfaces of the Rrp41/45 dimer; in particular, it interacts with the C-terminal tail of Rrp45

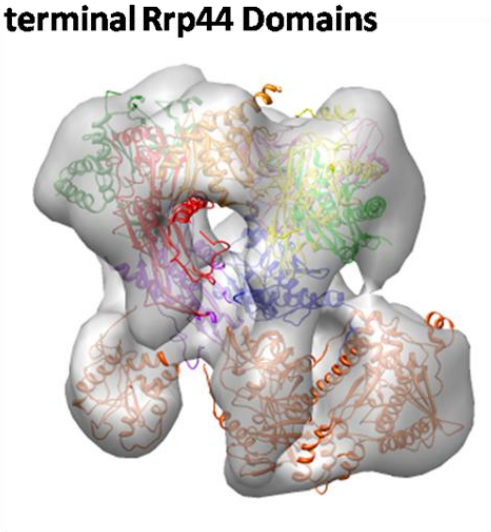
Figure 2-10. Crystal structure of the Rrp44/41/45 ternary complex and docking of the structure into the yeast exosome EM reconstruction. A) Crystal structure of the Rrp44 25-1001/Rrp41/Rrp45 structure (pdb 2wp8). Rrp41 and Rrp45 are in solid colors, while Rrp44 is color coded by domain. Regions of interaction between Rrp44 and Rrp41 and Rrp45 occur mainly between loops and are indicated by circles. This figure was created in PYMOL (*19*). B) The crystal structures of the human exosome (*1*) (pdb 2nn6), after omitting subunits Rrp41, Rrp45, and Csl4, and the *S. cerevisiae* Rrp41/Rrp45/Rrp44 25-1001 ternary complex (*18*) (pdb 2wp8) were docked into the EM reconstruction of the yeast exosome (*6*) using the program UCSF Chimera (*20*). Rrp44 is in solid red-orange, while Rrp41 and Rrp45 are colored following the same color scheme as Fig. A. On the left, the ternary complex structure was docked in its entirety. On the right, the N and C-terminal regions of Rrp44 were separated before docking.



Whole Rrp44/41/45 Complex Docking



Docking After Splitting N and C terminal Rrp44 Domains



(residues 293-303) by means of hydrophobic and polar contacts (*18*). We initially identified several stable ternary complexes of Rrp44 to Rrp41 and Rrp45, and our studies produced an N-terminal truncation of Rrp44 (Δ17) that was very similar to that used by the Conti group (Δ25). We were also correct that the N-terminal PIN domain region of Rrp44 was necessary and sufficient for attachment to the exosome; however, C-terminal regions of interaction within Rrp44 were more difficult to predict. Ultimately, these C-terminal interactions may have provided more crystal contacts, which could have aided in producing better diffracting crystals and a high resolution structure.

Another reason why crystallization could have been impeded was the manner in which the ternary complex was assembled. In the study published by the Conti group (*18*), the complex was assembled by mixing Rrp44 with the Rrp41/45 complex, further purifying it, and pooling fractions containing the complex for crystallization. The initial manner in which we purified the Rrp44 1-270/Rrp41/45 complex followed similar methods, and crystallized readily. However, in order to reduce the amount of labor for further studies, purification of subsequent complexes was completed using co-expression instead of mixing, but none of these complexes crystallized. Therefore, it is possible that co-expression of the ternary complex resulted in poorly assembled or sub-stoichiometric complexes, which impeded crystallization. During co-expression, it was noticed that the Rrp41/45 complex always had a much higher level of expression than that of the Rrp44 protein, and excess Rrp41/45 was always collected during second step MonoQ purification. We noticed a similar situation occurred when purifying recombinant exosomes using a co-expression technique (see Chapter 3). Exosomes expressed in this manner also had an over-abundance of Rrp41/45 before MonoQ separation, and were poorly assembled when viewed under EM (*21, 22*). Therefore, it is possible that the Rrp44/Rrp41/Rrp45 ternary complexes

purified using co-expression could have been improperly assembled as well.

2.8. Relevance of the ternary complex structure in the context of the intact exosome

While the Rrp44/Rrp41/Rrp45 ternary complex provides insight into how Rrp44 associates with two subunits of the exosome core, it is important to know whether this structure is relevant within the physiological context of the complete exosome core. In order to gain insight into how Rrp44 associates with the whole exosome, the crystal structure of the Rrp44/Rrp41/Rrp45 ternary complex along with the structure of the human exosome core (after deleting subunits Rrp41, Rrp45, and Csl4) was docked into the EM reconstruction of the Rrp44 bound yeast exosome (*6*). As shown in Fig. 2-10B, and as reported previously (*6*) the human exosome core docks into the yeast exosome EM reconstruction quite well, despite the fact that it is from a different organism. The N-terminal PIN region of Rrp44 partially docks into the EM reconstruction, while the C-terminal body region of Rrp44 docks rather poorly into the density from the EM reconstruction (Fig. 2-10, left). A much better docking result can be obtained when separating the N and C-terminal regions of Rrp44 and docking them separately (Fig. 2-10, right). The EM reconstruction of the Rrp44 exosome suggested that the PIN domain of Rrp44 is separated from the C-terminal region by about 20 Å, while the crystal structure of the ternary complex suggests a much closer proximity. Therefore, it is possible that the conformation of the PIN domain seen in the crystal structure is the result of a crystal packing artifact and may not accurately reflect how Rrp44 associates with the core exosome. Further crystallographic studies of a complete Rrp44 bound exosome would resolve the disparities between the present crystal structures and EM reconstruction.

The published structure of the Rrp44/Rrp45/Rrp41 ternary complex also raised questions about how RNA interacts with Rrp44 when this subunit is bound to the exosome. A prior structure of a truncated Rrp44 without the PIN domain solved in complex with single stranded, polyadenine RNA (*12*) suggested that the RNA made most of its contacts in the RNB domain of Rrp44. In the RNA bound structure of Rrp44ΔPIN, amino acids 696-719 fold into a short beta sheet and loop that is involved in RNA binding (Fig. 2-11, *left*) (*12*). However, in the Rrp44/Rrp41/Rrp45 structure, this same region is blocked by a conformation change in Rrp44, in which residues 696-719 fold into a short helix (Fig. 2-11, *right*), and it was suggested that this inhibited conformation could be stabilized upon Rrp44 binding to the exosome since residue R303 from Rrp45 could electrostatically repel residue K702 from Rrp44 which is involved in RNA binding (*18*). Biochemical results also suggest that a single stranded RNA of 31-33 nucleotides is channeled through the exosome core and then degraded by Rrp44 (*18*) (discussed in greater detail in Chapter 3), in a manner reminiscent of the archaeal exosome, and suggest that the previous reported RNA access pathway may thus be irrelevant for activity when Rrp44 is bound to the yeast exosome. However, such as recruitment pathway could be relevant for Rrp44 like enzymes in humans, such as Dis3L2, which lacks a PIN domain and presumably does not bind to the exosome. The RNA access route depicted in the Rrp44ΔPIN structure could also be relevant for a "direct access pathway" of RNA degradation by the exosome, which was initially proposed based on structural comparison of Rrp44 with the homologous bacterial enzyme RNase II (*23, 24*). This direct access pathway will be studied in greater biochemical detail through the use of biochemical assays in the next chapter.

2.9. The RNA binding channel is blocked in the Rrp44ΔPIN structure as well.

The assumption that the exosome regulated Rrp44 by stabilizing an inhibited

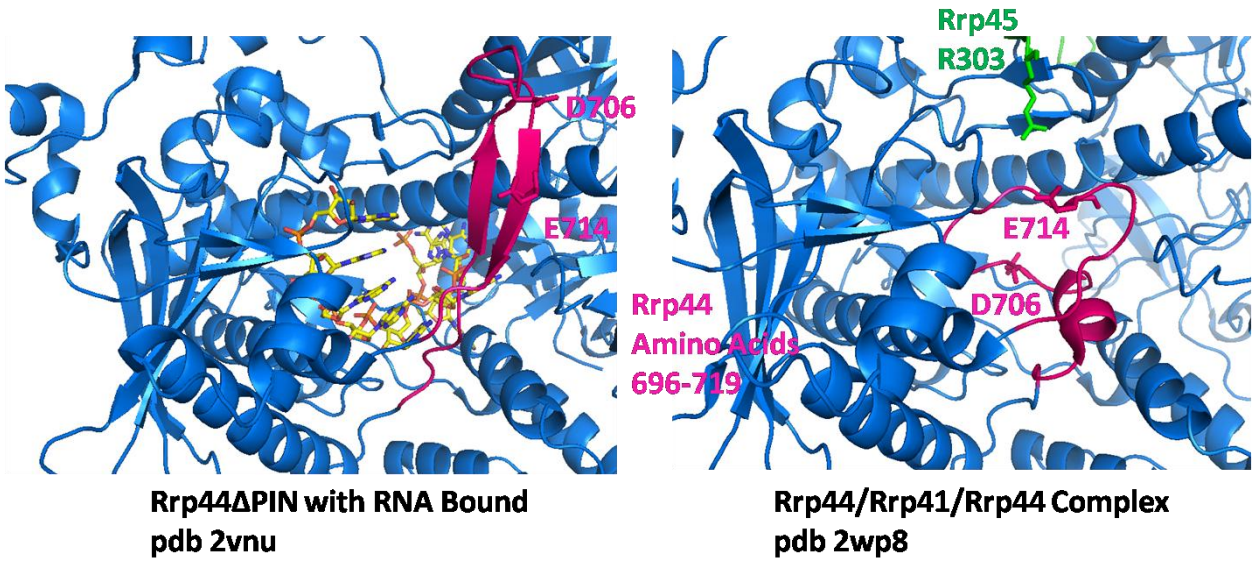


Figure 2-11. Comparison of the active sites of the Rrp44/Rrp41/Rrp45 structure with that of the Rrp44ΔPIN RNA bound structure. In the Rrp44ΔPIN RNA bound structure (left), amino acids 696-719 form a short beta sheet and loop involved in RNA binding in the active site. In the Rrp44/Rrp41/Rrp45 structure (right), the active site is blocked by a conformational change that causes this region to fold into a short alpha helix and loop. Figure illustrated with PYMOL software (*19*).

conformation that blocks its active site upon binding, as suggested by the Conti group (*18*), is an interesting yet untested hypothesis. Since the Rrp44 region 696-719 occluded the RNA access route deep within the Rrp44 active site channel (*18*), it would be difficult to determine how RNA would be degraded at all when Rrp44 assumed such a conformation. Since no apo structure of Rrp44ΔPIN had been solved, it is possible that the "open state" of the active site channel of Rrp44 was caused by binding to RNA. Moreover, little evidence was present to suggest that the inhibited conformation of Rrp44 could be stabilized by Rrp45, since the conformation of this protein could change upon RNA binding as well. In order to clear up this discrepancy, we set about to solve the structure of the Rrp44ΔPIN apo structure.

2.9.1. Methods

Cloning, expression, and protein purification

N-terminal hexahistidine tagged Rrp44 242-1001 exonuclease mutant (D551N) was amplified from a plasmid containing this mutation, restricted with NcoI and XhoI enzymes, and ligated into a pET19b vector. The protein was expressed and purified by Ni-NTA chromatography using methods similar to those described previously for other Rrp44 proteins, and underwent a final MonoQ purification step. Protein was then dialyzed overnight into 10 mM Tris pH 8.0, 150 mM NaCl, 1 mM DTT, and 2.5 mM $MgCl_2$. Protein was concentrated to 30.4 mg/mL, snap frozen, and stored at -80 °C until use.

Crystallization

Protein at a concentration of 15.2 mg/mL was crystallized using hanging drop wells over a reservoir of 50 mM MES pH 6.0, 150 mM NaCl, 1 mM DTT, 2.5 mM $MgCl_2$, and 12% PEG 400. Crystals appeared as thin plates (Fig. 2-12A).

Data Collection and Structure Determination

Protein crystal diffraction data was collected at CHESS beamline A1. Phase information was obtained by the molecular replacement using the Rrp44ΔPIN RNA bound structure (pdb 2vnu) (*12*) as a model. The model was fit into the electron density using coot (*25*), and underwent rigid body refinement in CNS (*26*), followed by further refinement in REFMAC (*27*) from the CCP4 software suite (*28*). Crystallization statistics are found in Table 2-2.

2.9.2. Results

In the structure of Rrp44ΔPIN apo, four molecules (A, B, C, and D) were found in the asymmetric unit (Fig. 2-12B), and aligned with a root mean squared deviation of 0.65 – 0.69 Å (Fig. 2-12C). Although the structure of the Rrp44ΔPIN apo was of a lower resolution (3.2 Å) than the RNA bound form (2.3 Å), an overall similar domain organization could be determined (Fig. 2-12D) and agreed well with that of both the RNA bound form (pdb 2vnu) as well as the Rrp44/41/45 structure (pdb 2wp8).

Three (molecules A, B, and D) out of four molecules contained continuous electron density around the active site channel (amino acids 696-719), while poorer density was present for this region in molecule C (Fig. 2-14A-B). This region folds into a short alpha helix and loop that encompasses the RNA binding region of the active site, and assumes a conformation almost identical to that found in the structure of the Rrp44/Rrp41/Rrp45 ternary complex (Fig. 2-14 C). Therefore, it appears that the active site is inhibited in the apo form of Rrp44, and that binding to RNA causes a conformational change that allows for RNA entry and subsequent degradation. Further studies will need to be conducted to determine whether a conformational change occurs

Table 2-2. Crystallography Statistics of Rrp44ΔPIN Apo

Parameter	Rrp44ΔPIN Apo
Beamline	MacCHESS A1
Space group	$P6_4$
Unit cell parameters (Å)	a = b = 145.50, c=309.97
Wavelength (Å)	0.9770
Resolution range (Å)	50 – 3.2 (3.26 3.20)
Measured reflections	174736
R_{work} (%)	26.02
R_{free} (%)	30.02
Unique reflections	55433
Redundancy	3.2 (2.0)
Completeness	90.6 (80.7)
Average I/σ(I)	9.08 (1.58)
Rmerge (%)	11.8 (30.5)
Molecules per Asymetric unit	4

in the exosome subunits upon RNA binding as well.

Figure 12-12. Crystallization and structure of Rrp44ΔPIN apo. A) Plate crystals of the proteins. B) Asymmetric unit of Rrp44ΔPIN apo containing four molecules. C) Alignment of the four molecules found within the asymmetric unit. D) Domain organization of one monomer (molecule A) agrees well with other published Rrp44 structures. The color scheme of the domains is the same as that in figure 1-3. Figures B and C are courtesy of Ki Hyun Nam.

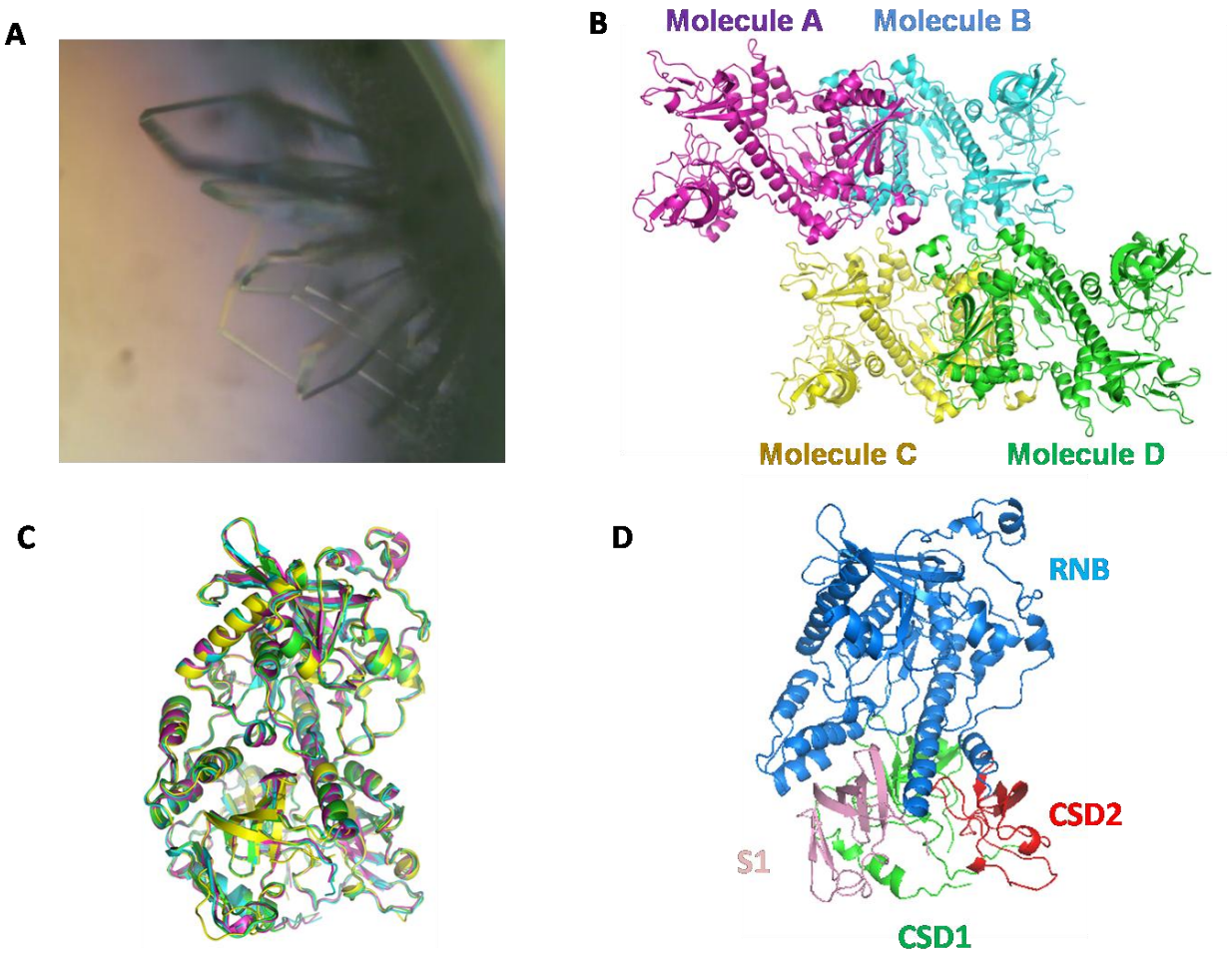
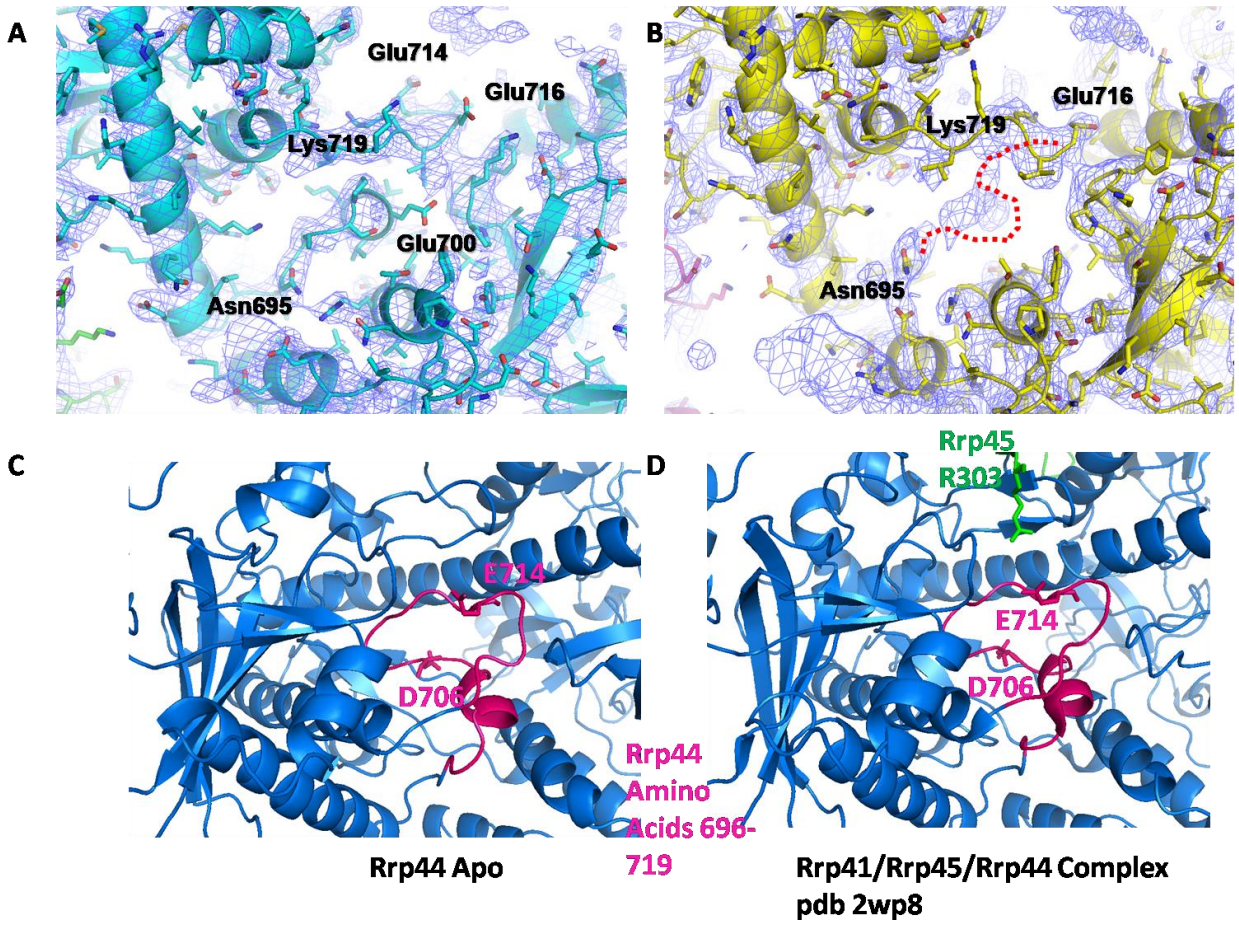

A
B
Molecule A
Molecule B
Molecule C
Molecule D
C
D
RNB
CSD2
S1
CSD1

Figure 12-13. The active site of Rrp44 apo. A) Electron density for molecule B indicating good density for the region of residues 696-719 and B) poorer density for this region found in molecule C. C) The active site of Rrp44ΔPIN apo assumes the same conformation as that of the ternary complex between Rrp44/Rrp41/Rrp45 (pdb 2wp8). This figure was created with PYMOL (*19*). Figures A and B are courtesy of Ki Hyun Nam.


A
Glu714
Glu716
Lys719
Glu700
Asn695
B
Glu716
Lys719
Asn695
C
E714
D706
Rrp44 Apo
D
Rrp45
R303
E714
D706
Rrp44
Amino
Acids 696-
719
Rrp41/Rrp45/Rrp44 Complex
pdb 2wp8

References

1. Liu, Q., Greimann, J. C., and Lima, C. D. (2006) Reconstitution, activities, and structure of the eukaryotic RNA exosome, *Cell 127*, 1223-1237.
2. Chen, C. Y., Gherzi, R., Ong, S. E., Chan, E. L., Raijmakers, R., Pruijn, G. J., Stoecklin, G., Moroni, C., Mann, M., and Karin, M. (2001) AU binding proteins recruit the exosome to degrade ARE-containing mRNAs, *Cell 107*, 451-464.
3. Staals, R. H., Bronkhorst, A. W., Schilders, G., Slomovic, S., Schuster, G., Heck, A. J., Raijmakers, R., and Pruijn, G. J. (2010) Dis3-like 1: a novel exoribonuclease associated with the human exosome, *The EMBO journal 29*, 2358-2367.
4. Tomecki, R., Kristiansen, M. S., Lykke-Andersen, S., Chlebowski, A., Larsen, K. M., Szczesny, R. J., Drazkowska, K., Pastula, A., Andersen, J. S., Stepien, P. P., Dziembowski, A., and Jensen, T. H. (2010) The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L, *The EMBO journal 29*, 2342-2357.
5. Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M., and Tollervey, D. (1997) The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-->5' exoribonucleases, *Cell 91*, 457-466.
6. Wang, H. W., Wang, J., Ding, F., Callahan, K., Bratkowski, M. A., Butler, J. S., Nogales, E., and Ke, A. (2007) Architecture of the yeast Rrp44 exosome complex suggests routes of RNA recruitment for 3' end processing, *Proceedings of the National Academy of Sciences of the United States of America 104*, 16844-16849.
7. Dziembowski, A., Lorentzen, E., Conti, E., and Seraphin, B. (2007) A single subunit, Dis3, is essentially responsible for yeast exosome core activity, *Nature structural & molecular biology 14*, 15-22.
8. Heckman, K. L., and Pease, L. R. (2007) Gene splicing and mutagenesis by PCR-driven overlap extension, *Nature protocols 2*, 924-932.
9. Aslanidis, C., and de Jong, P. J. (1990) Ligation-independent cloning of PCR products (LIC-PCR), *Nucleic acids research 18*, 6069-6074.
10. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Analytical biochemistry 72*, 248-254.
11. Zor, T., and Selinger, Z. (1996) Linearization of the Bradford protein assay increases its sensitivity: theoretical and experimental studies, *Analytical biochemistry 236*, 302-308.
12. Lorentzen, E., Basquin, J., Tomecki, R., Dziembowski, A., and Conti, E. (2008) Structure of the active subunit of the yeast exosome core, Rrp44: diverse modes of substrate recruitment in the RNase II nuclease family, *Molecular cell 29*, 717-728.
13. Ishida, T., and Kinoshita, K. (2007) PrDOS: prediction of disordered protein regions from amino acid sequence, *Nucleic acids research 35*, W460-464.
14. Ke, A., and Wolberger, C. (2003) Insights into binding cooperativity of MATa1/MATalpha2 from the crystal structure of a MATa1 homeodomain-maltose binding protein chimera, *Protein science : a publication of the Protein Society 12*, 306-312.
15. Shiomi, T., Fukushima, K., Suzuki, N., Nakashima, N., Noguchi, E., and Nishimoto, T. (1998) Human dis3p, which binds to either GTP- or GDP-Ran, complements Saccharomyces cerevisiae dis3, *Journal of biochemistry 123*, 883-890.

16. Korbie, D. J., and Mattick, J. S. (2008) Touchdown PCR for increased specificity and sensitivity in PCR amplification, *Nature protocols 3*, 1452-1456.
17. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic acids research 22*, 4673-4680.
18. Bonneau, F., Basquin, J., Ebert, J., Lorentzen, E., and Conti, E. (2009) The yeast exosome functions as a macromolecular cage to channel RNA substrates for degradation, *Cell 139*, 547-559.
19. The PyMOL Molecular Graphics System, Version 1.5.0.1 Schrödinger, LLC.
20. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis, *Journal of computational chemistry 25*, 1605-1612.
21. Liu, X., and Wang, H. W. (2010) Preliminary Electron Microscopy Studies of Recombinant Exosomes, Yale University School of Medicine.
22. Lee, G., and Ha, T. (2010) Preliminary Single Molecules Studies of Recombinant Exosomes, University of Illinois.
23. Frazao, C., McVey, C. E., Amblar, M., Barbas, A., Vonrhein, C., Arraiano, C. M., and Carrondo, M. A. (2006) Unravelling the dynamics of RNA degradation by ribonuclease II and its RNA-bound complex, *Nature 443*, 110-114.
24. Zuo, Y., Vincent, H. A., Zhang, J., Wang, Y., Deutscher, M. P., and Malhotra, A. (2006) Structural basis for processivity and single-strand specificity of RNase II, *Molecular cell 24*, 149-156.
25. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta crystallographica. Section D, Biological crystallography 60*, 2126-2132.
26. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta crystallographica. Section D, Biological crystallography 54*, 905-921.
27. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta crystallographica. Section D, Biological crystallography 53*, 240-255.
28. (1994) The CCP4 suite: programs for protein crystallography, *Acta crystallographica. Section D, Biological crystallography 50*, 760-763.

Chapter 3: Biochemical studies of Rrp44 and the Rrp44 bound exosome

Abstract

This chapter entails the use of yeast exosomes for activity assays with structured RNA substrates containing with various 3' end tail lengths. We first detail the purification of wild type yeast exosomes using the TAP method and the reconstitution of core exosomes with mutated Rrp44. Rrp44 and Rrp44 containing exosomes were then used in both exo- and endo- nuclease assays with both 3' end extended forms of both the Hepatitis Delta Virus Ribozyme as well as yeast initiator methionine tRNA. We find that a 3' end tail length of ~8-9 nucleotides is necessary for exonuclease degradation by Rrp44, while degradation by the exosome is down-regulated and only occurs with a tail length of about 20 nucleotides. This tail length is ~10-13 nucleotides shorter than the nucleotide length required to span from the entrance of the exosome channel to the active site of Rrp44, indicating a direct access pathway to Rrp44 for degradation. Furthermore, we show that while the N-terminal endonuclease PIN domain of Rrp44 can degrade single stranded regions of structured RNA promiscuously, association with the exosome results in the restriction of degradation to the 3' end tail of the RNA. The PIN domain was also shown to contribute to RNA binding to structured substrates. The fact that the 3' end tail of the RNA can be degraded by both exo and endonuclease activities place the RNA in a position to interact with both the N- and C-terminal regions of Rrp44. When taken together with the length requirement of RNA degraded and collaborative EM reconstruction and single molecule studies, we propose that RNA can access Rrp44 when bound to the exosome by using an RNA entrance pathway reminiscent of RNase II.

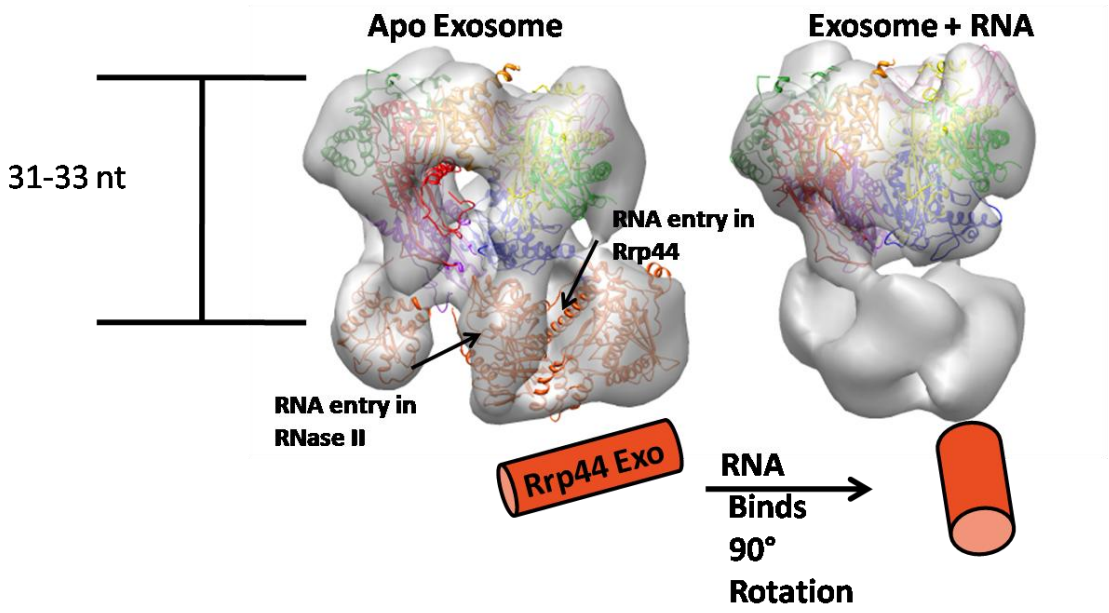
3.1. Introduction

The EM model of the yeast exosome (*1*) suggested two possible routes of RNA recruitment to the exosome: 1) the through exosome route, in which RNA passes through the core exosome channel and is then feed into Rrp44 for degradation, and 2) the direct access route in which RNA is directly recruited to Rrp44 while avoiding the exosome channel. At this time, evidence from the Conti group indicates that single stranded RNA is degraded via the through exosome route (*2*). The research group made mutations in the entry and exit sites to the RNA channel in the exosome core and showed that degradation of single stranded RNA was inhibited (*2*). However, the experiment was somewhat biased since it was conducted by pre-binding substrates with purified recombinant core exosomes, thus favoring exosome channeling, before adding Rrp44 for degradation assays. The report indicated that a length of 31-33 nucleotides was required to reach from the entrance of the exosome to the active site of Rrp44 for degradation (Fig. 3-1A, *left*) (*2*). While this model explains a probable route for single stranded RNA degradation, it does not explain how more structured substrates can be degraded by the exosome.

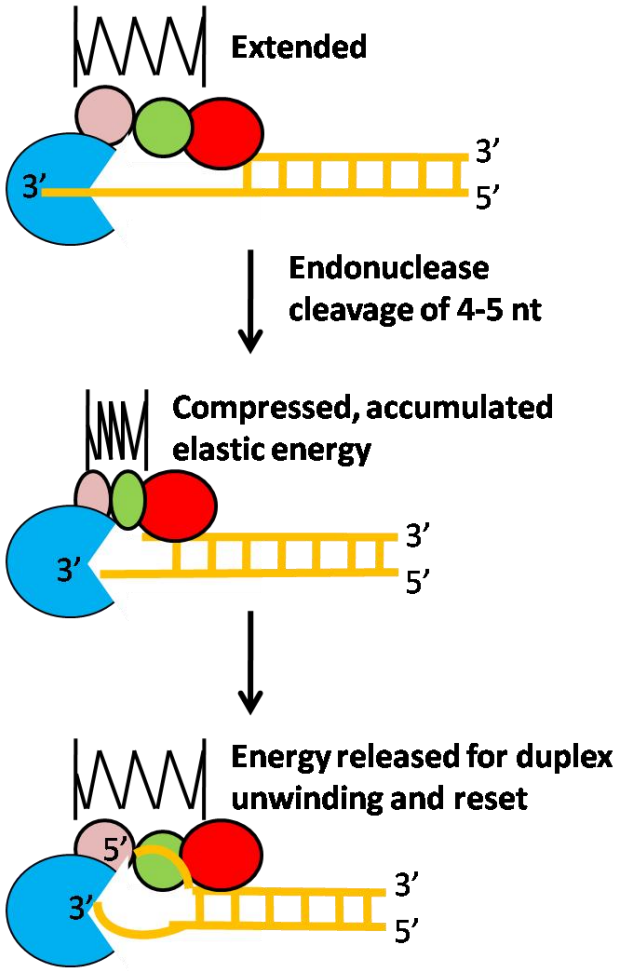
Although the through exosome route appears to be the favored pathway for degradation of single stranded RNA, little work has been conducted to test the direct access route. In particular, the direct access route is an intriguing pathway for structured RNA substrates. Evidence to support the existence of this route includes the fact that in the co-crystal structure of Rrp44ΔPIN bound to RNA, the 5' end of the RNA is pointed towards solvent (Fig. 3-1A, *left*), indicating that this pathway is not continuous with the exosome core (*3*). Furthermore, a second pathway into the active site of Rrp44 could occur which is more reminiscent of the *E. coli* enzyme RNase II (*4*) (Fig. 3-1A, *left*). Also, the PIN domain

Figure 3-1. Mechanisms of RNA processing and unwinding by Rrp44. A) Reported pathway of Rrp44 mediated degradation by the exosome, shown as a dashed line. A length of 31-33 nucleotides is reported to reach from the entry site of the exosome core to the active site of Rrp44 (*2*). Depicted is the docking of the crystal structures of the Rrp44/Rrp41/Rrp45 complex (pdb 2wp8) and the human exosome, with Rrp41, Rrp45, and Csl4 subunits removed, (pdb 2nn6) into the EM reconstruction of the Rrp44 bound exosome (gray envelope, EM Databank ID 1438). Docking was performed with UCSF Chimera (*5*). B) Crystal structure of Rrp44ΔPIN (light shades, pdb 2vnu) aligned with RNase II (dark shades, pdb 2ixi) with domains color coded and RNA access routes depicted. C) Proposed mechanism of unwinding of duplex RNA by Rrp44. Circles represent domains of Rrp44, and are colored according to the scheme in B. Depicted above the CSD2 and S1 domains is a cartoon spring to represent how these domains compress and release during RNA degradation. This figure is based on data from a single molecule FRET study (*6*).

A


Apo Exosome
Exosome + RNA
31-33 nt
RNA entry in Rrp44
RNA entry in RNase II
Rrp44 Exo
RNA Binds 90° Rotation

B


Extended
3′
5′
3′
Endonuclease cleavage of 4-5 nt
Compressed, accumulated elastic energy
3′
5′
3′
Energy released for duplex unwinding and reset
5′
3′
3′
5′

of Rrp44 contains endonuclease activity (*7-9*). The active site of the PIN domain of Rrp44 is also solvent exposed (Fig. 3-1A, *left*), so it does not appear to be in a position to interact with RNA emerging from the exosome core (*2*). Recent preliminary negative stain EM reconstruction efforts conducted in collaboration between the Ke and Wang groups indicate that the C-terminal region of Rrp44 undergoes a ~90° conformational change upon structured RNA binding, bringing it closer to the PIN domain (*10*) (Fig. 3-1A, *right*). Since the conformational change occurs only in Rrp44 and not any subunits of the exosome core subunits, it suggests that RNA is directly binding to Rrp44 in a direct access manner, and RNA binding would likely open up the active site of Rrp44 which was blocked in the Rrp44ΔPIN apo, and Rrp44/Rrp41/Rrp45 (*2*) structures. Moreover, since the conformational change occurs within a region in close proximity of the RNase II like direct access degradation, this route is favored over the solvent exposed pathway into the active site of Rrp44. Further cryo-EM studies will need to be conducted to visualize the position of the RNA and to confirm the veracity of this hypothesis.

The ability of the exosome to degrade structured RNAs that are too large to pass through the core channel is reconciled by evidence that these substrates can be unwound by helicase containing complexes such as the nuclear TRAMP complex (*11-13*). Numerous studies have been conducted to show that the nuclear exosome can be activated by the TRAMP complex to degrade RNA (*11-13*), which has also be reported to "activate" the exosome. The majority of studies conducted with the TRAMP complex, however, do not differentiate between the degradation activity of Rrp6 and Rrp44 in exosome preparations. Furthermore, more recent reports indicate that the activity of Rrp6 is most likely activated by TRAMP, but the activity of Rrp44 is not (*14, 15*). Few degradation studies have been conducted using the Rrp44 bound exosome in the absence of other helicase factors such as the TRAMP complex. However, one

study did indicate that Rrp44 could partially, yet selectively, degrade an initiator methionine tRNA that was hypomethylated on nucleotide A58, out of a pool of RNA (*13*). Although it is possible that the tRNA was partially unfolded due to the missing methylation site, it is unlikely that it would have had an exposed 3' end of 31-33 nucleotides that could have been fed into the exosome channel. This suggests that partially folded tRNA may be able to access the Rrp44 bound exosome via the direct access pathway.

Although the nuclear TRAMP complex and cytoplasmic SKI complex both contain helicase activities which could activate substrates for exosome mediated degradation, evidence indicates that the Rrp44 bound exosome contains an intrinsic helicase activity itself (*3*). The Conti group previously showed that duplex substrates with a 3' end of at least 7 nucleotides could be unwound and degraded by the Rrp44 alone (*3*), while duplexes with a 3' end of 35 nucleotides could be unwound and degraded by the Rrp44 containing exosome (*2*). Furthermore, a study conducted by the Ha lab (University of Illinois—Urbana Champagin) in collaboration with the Ke group indicted that Rrp44 first degrades a short single stranded overhang and then unwinds the duplex in discrete step sizes of 4-5 bp/step (*6*). It was proposed that about seven nucleotides of the 3' overhang of the substrate are bound in the active site of Rrp44 during degradation, leaving the single/double stranded RNA junction in proximity to CSD1 and CSD2. The CSD domains abut this junction and are positioned to act as "wedges" while the RNB domain degrades 4-5 additional nucleotides. This causes tension in the 3' single stranded tail and accumulates elastic energy that is utilized to unwind several base pairs of duplex RNA in a burst fashion, resulting in the movement of the CSD wedges (Fig. 3-1C). This cycle is continued until the double stranded RNA region is transversed. Thus, the data show that Rrp44 acts as an "elastic helicase" to unwind RNA using the hydrolysis of the single stranded overhang to drive

the reaction (*6*). If a similar mechanism exists while Rrp44 is bound to the exosome core, this helicase ability may be relevant to degrade structured substrates by the direct access pathway without the use of either the TRAMP or SKI complexes.

In order to more thoroughly determine whether a direct access pathway of degradation to the exosome occurs, we tested a number of structured substrates with Rrp44 and Rrp44-bound exosomes in degradation assays. The main substrates tested were the Hepatitis Delta Virus ribozyme and hypo-modified initiator methionine tRNA. Substrates used had a structured core followed by a 3' end overhang extension of various lengths. The 3' end tail was added to serve as a "molecular ruler" to determine the length of single stranded RNA that was required to be actively degraded by either Rrp44 or the exosome, to demonstrate whether a direct access route to Rrp44 occurred, which would be indicated by degradation of a 3' end region that spanned a distance of less than 31-33 nucleotides—the reported length from the exosome ring to the active site of Rrp44 (*2*). Identification of such as pathway will more holistically characterize the catalytic mechanism of the eukaryotic exosome.

3.2. Purification of recombinant Rrp44 bound exosomes via cotransformation

In order to study the activity of the exosome, it is essential to purify stable Rrp44 bound exosomes for biochemical assays. The eukaryotic exosome is composed of three heterodimers (Rrp41/Rrp45, Rrp43/Rrp46, and Rrp42/Mtr3) and three RNA binding "capping" proteins (Rrp4, Csl4, and Rrp40), which bind at an interface created between adjacent heteordimers (*16*). Our results, as well as those from others (*17, 18*) indicate that while the heterodimers purify well when expressed with their respective binding partner, they express poorly alone. It has been

suggested that co-expression of proteins with their binding partner buries hydrophobic residues that cause the proteins to be insoluble when expressed alone (*17*). In order to avoid insolubility issues during purification, we initially attempted a co-expression procedure. Previous published methods detailed the assembly of exosomes by mixing the three individually purified heterodimers with the three capping proteins and then purifying the complex (*17*). The goal of our procedure, however, is to co-express the exosome in as few steps as possible, and is described below.

3.2.1 Methods

Cloning

Three polycistronic vectors containing the majority of the exosome genes were used for the purpose of co-expression, and were constructed by Ben Clarke, a former rotation student in the lab, and Fran Ding. Polycistronic insert regions of exosome genes *Rrp42-Mtr3-Rrp43; Rrp45-Rrp41-Rrp46;* and *Rrp4-Rrp40* were amplified from genomic DNA, and cistrons were created using overlap extension PCR (*18*), and inserted into vectors containing N-terminal hexahisidine tags for purification as summarized in Fig. 3-2. The *Rrp42-Mtr3-Rrp43* insert was ligated into a pET23a vector containing ampicillin resistance using 5' KpnI and 3' BamHI restriction sites. The *Rrp45-Rrp41-Rrp46* pACYC184 vector was prepared as discussed previously (Chapter 2.2.1). The *Rrp4-Rrp40* insert was cloned into a pLIC vector containing kanamycin resistance. All plasmids were transformed into DH5α cells, selected with the appropriate antibiotic, mini-prepped, and sequenced to confirm the presence of the correct insert. All three plasmids were then co-transformed into BL21* cells and grown on ampicillin, tetracycline, and kanamycin

triple selection plates.

The remaining exosome plasmids (Rrp44 and Csl4) were constructed by Fran Ding. The gene for full length wildtype Rrp44 was inserted into a hexahistidine pLIC vector with a kanamycin selection marker, and Rrp44 mutants (D551N, D171N, or D551N-D171N) were created by site directed mutagenesis of the wildtype plasmid. Csl4 was inserted into a pET23a vector containing an N-terminal hexahistidine tag and ampicillin selection marker. Both plasmids were prepared using standard procedures and confirmed by DNA sequencing.

Protein Expression and Purification

Cultures containing recombinant exosomes were grown in 4.5 L of LB supplemented with concentrations of 5 µg/mL tetracycline, 100 µg/mL ampicillin, and 50 µg/mL kanamycin in a shaker at 37 °C and 200 rpm until reaching an OD_{600} of ~1. They were then induced with 1 mM IPTG and incubated overnight at 18 °C and 200 rpm. Cells were then pelleted and resuspended in Buffer A (20 mM Tris pH 7.5, 300 mM NaCl, 10 mM Imidazole, 5 mM betamercaptoethanol, 10% glycerol, and 1 mM PMSF) and were lysed by sonication. Purification was completed using Ni-NTA chromatography. Supernatant was applyed to the resin, washed with 10 column volumes of Buffer A containing 20 mM imidazole, and eluted with Buffer A containing 300 mM imidazole. The complex was then dialyzed into 20 mM Tris pH 7.5, 75 mM NaCl, 5 mM betamercaptoethanol, and 10% glycerol overnight, and run on MonoQ using the same buffer and eluted with an increasing NaCl gradient. Typical purifications of the recombinant exosome resulted in two nearly overlapping peaks (Fig. 3-3A). Closer analysis revealed that the first of these peaks contained exosomes lacking subunit Rrp43, while the second peak contained complete exosomes. Only peaks containing complete exosomes were pooled and concentrated.

Full length wild type Rrp44 and mutants as well as Csl4 were grown in 4.5 L of LB

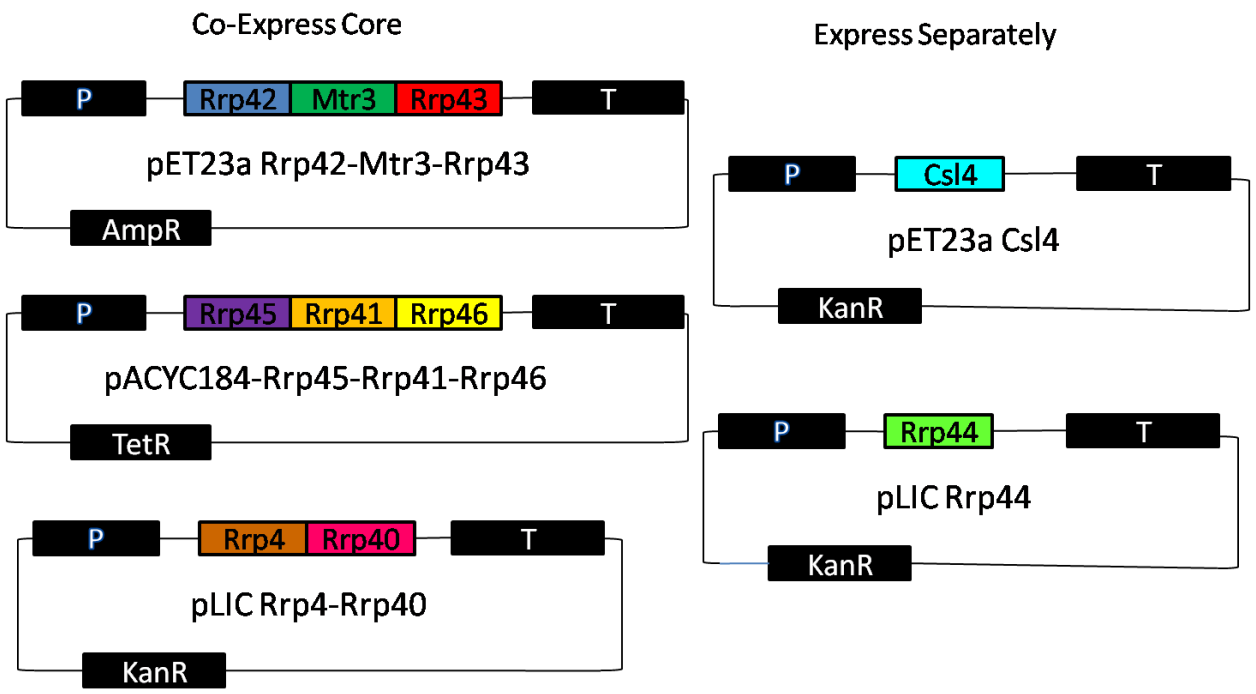


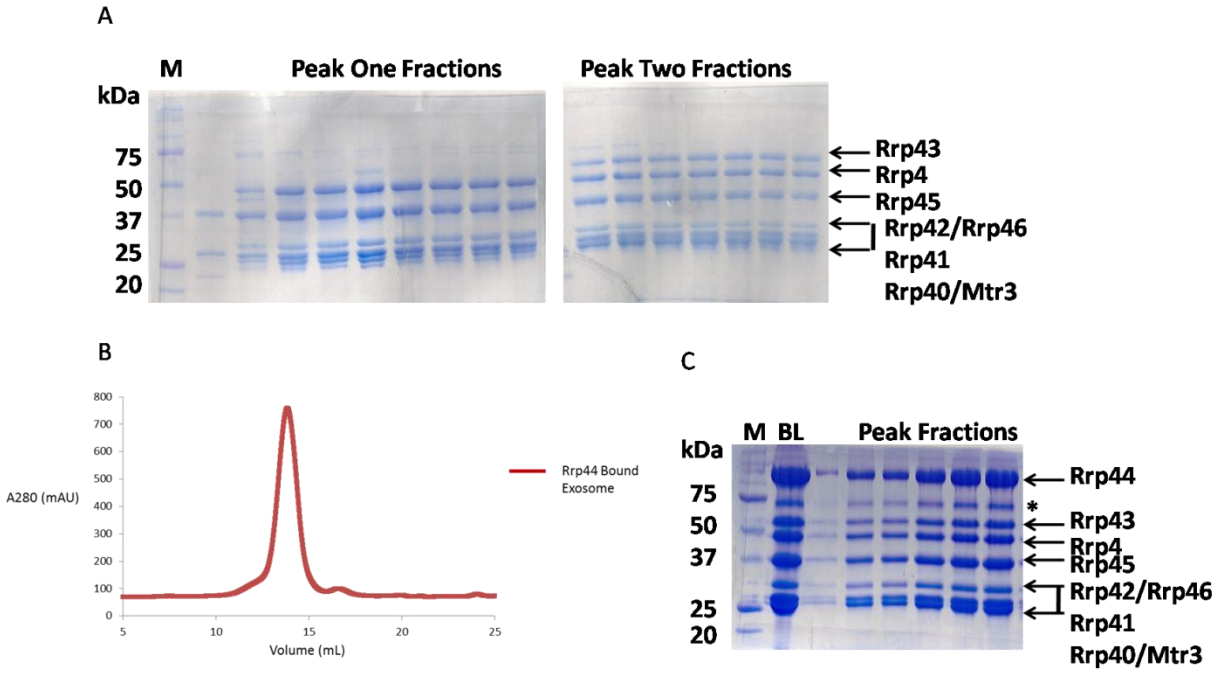
Figure 3-2. Summary of the co-expression strategy used to purify recombinant exosomes. Plasmids on the left were co-expressed under triple antibiotic selection, while those on the right were expressed separately and then mixed with the co-expressed proteins.

supplemented with the appropriate antibiotic, and expressed and purified in a similar fashion as for the exosome, and concentrated. The partially assembled exosome was then mixed in a 1:1.5:1 molar ratio with Csl4 and Rrp44 (wild type or mutant), incubated on ice, and separated on a Superose 6 column equilibrated with 20 mM Tris pH 8.0, 150 mM NaCl, 10% glycerol, and 1 mM DTT. The complex eluted as a single peak (Fig. 3-3B), which appeared to contain all exosome subunits (Fig. 3-3C).

3.2.2. Results

Recombinant exosomes were intended for EM analysis and biochemical assays. Despite appearing to assemble well during size exclusion chromatography, upon EM analysis exosomes were revealed to be poorly assembled (*19*). Subsequent collaborative single molecule studies also indicated that the recombinant exosomes lacked activity (*20*). Therefore, it appears that recombinant co-expression is ineffective in producing well assembled exosomes. It is likely that not all subunits express to the same extent, and thus some subunits may remain substoichiometric in the complex, which could cause poor assembly and lack of activity. In particular, difficulty occurs with producing a stoichiometric level of Rrp43. It was suggested that the binding partner of Rrp43, Rrp46, undergoes degradation during purification but still binds to Rrp43, yet does not associate into exosomes (*17*). This would explain why onlyhalf the yield of the exosomes contain Rrp43. However, the exact reason for poor assembly and lack of activity is unknown. Although our failure to produce quality recombinant exosomes sheds some doubt on other studies conducted with recombinantly purified exosomes (*2*, *16*), the problems may be limited to our co-expression procedure and thus may not reflect the quality of exosomes purified by other methods.

Fig 3-3. Purification of Recombinant Exosomes. A) MonoQ purification of recombinant exosomes resulted in two overlapping peaks, the first of which lacked subunit Rrp43. B) Size exclusion chromatography profile of Rrp44 containing recombinant exosomes run on a Superose 6 column. C) SDS-PAGE analysis of Rrp44 containing exosomes after size exclusion chromatography purification. The asterisk indicates a contaminating band that likely represents a truncation of Rrp44.


A
M
Peak One Fractions
Peak Two Fractions
kDa
75
50
37
25
20
Rrp43
Rrp4
Rrp45
Rrp42/Rrp46
Rrp41
Rrp40/Mtr3
B
A280 (mAU)
800
700
600
500
400
300
200
100
0
5
10
15
20
25
Volume (mL)
Rrp44 Bound Exosome
C
M
BL
Peak Fractions
kDa
75
50
37
25
20
Rrp44
*
Rrp43
Rrp4
Rrp45
Rrp42/Rrp46
Rrp41
Rrp40/Mtr3

3.3. TAP purification of yeast exosomes

Since recombinant purification methods were unsuccessful, we instead purified exosomes from yeast cells using a Tandem Affinity Purificaiton (TAP) procedure (*21*), performed according to the following methods.

3.3.1. Methods

Yeast Culture Growth

A *S. cerevisiae* yeast strain containing a TAP tag on Rrp46 and a deletion for the nuclear exosome cofactor Rrp6 was kindly provided by J. Scott Butler of the University of Rochester Medical Center. A glycerol stock of the strain was streaked onto a YPD plate and grown overnight at 30 °C. A single colony was picked from the plate and inoculated into 200 mL of YPD media (10 g yeast extract, 20 g bacto-peptone, and 50 mg of adenine sulfate per liter) supplemented with 10 µg/mL of tetracycline, and grown for about 48 hours at 30 °C with shaking at 200 rpm. An amount of 16 mL of starter culture was then diluted into a total of 12 L of YPD media of yeast, supplemented with 4% glucose, and then grown at 30 °C for about 20 hours, or until the OD_{600} reached 6-10. At this point, the cultures were supplemented with 150 mL of 40% glucose and 37.5 mL of 4 g/L adenine sulfate. Flasks were shaken for another 6 hours until reaching a saturating OD_{600} of about 12-13 and then pelleted by centrifugation. A typical yield was about 450 g of cell paste. Cells were resuspended in 5X Lysis Buffer (250 mM Tris-HCl pH 7.38, 150 mM NaCl, 5 mM EDTA, 50% glycerol, 5% DMSO, and 50 µM $ZnCl_2$), supplemented with 5 mL of 100X complete protease inhibitors (17 mg PMSF, 28.5 µg leupeptin, 137 µg pepstatin A, and 33 mg benezamidine per mL dissolved in 100% ethanol), and flash

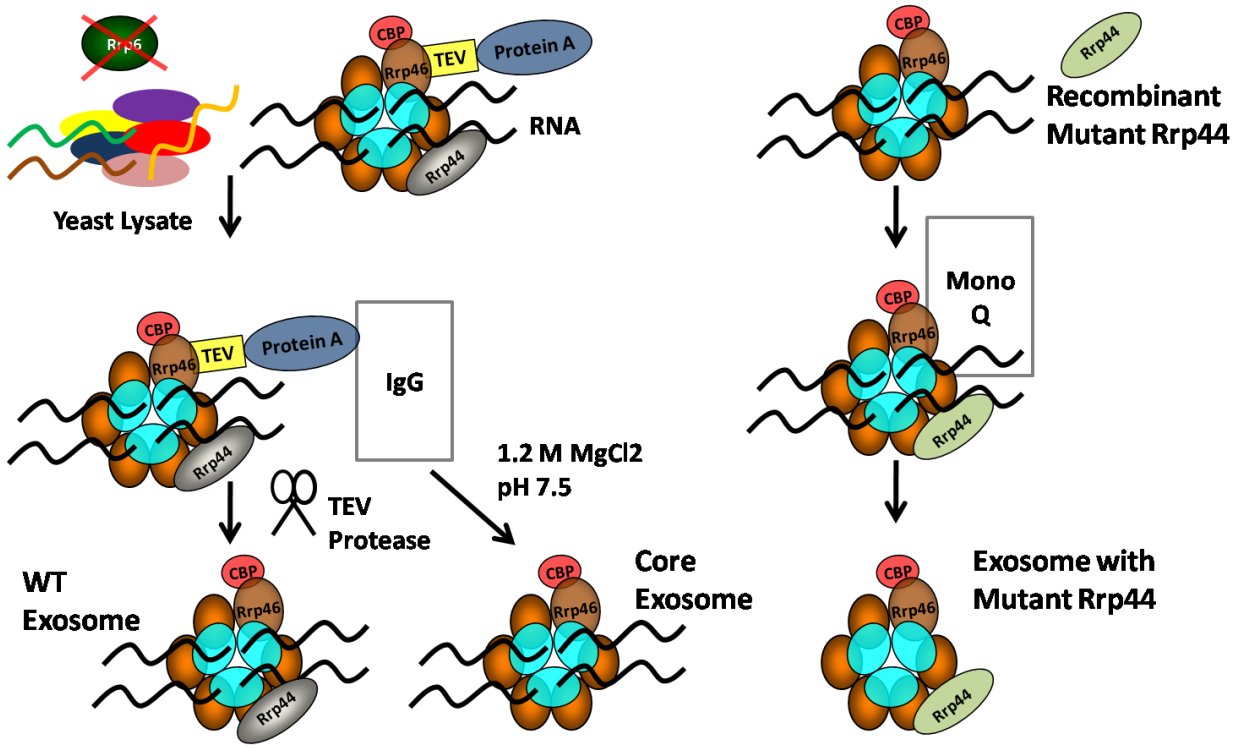


Figure 3-4. Cartoon representation of the TAP Purification of the Yeast Exosome. The straightforward procedure on the far left was used to purify exosomes containing wildtype Rrp44, while washing with 1.2 M $MgCl_2$ and subsequent application of mutant recombinant Rrp44 was used to purify mutant exsomes. All exosomes were polished with a MonoQ step, which removed contaminating proteins and all RNA (depicted as wavy lines).

frozen in 50 mL aliquots for storage at -80 °C.

TAP purification

For purification, cell pellets were thawed and solid NaCl was added to a final concentration of 150 mM. Cells were also supplemented with 5 mL of 100X protease inhibitors and 500 µL of 1 mg/mL DNase I solution. Cell disruption was completed by beating cells with silicon beads in two bead beaters for 1 hour at 4 °C. Bead beaters were submerged in an ice bath created by mixing ice with drive-way salt (IceMelt) to maintain a temperature of -10 °C to prevent protein denaturation from excessive heat released during cell breakage. Beads were removed by filtration, and lysate was centrifuged in an ultracentrifuge at 35,000 rpm for 1 hour. The TAP purification of all exosomes was completed as represented in Fig. 3-4. The supernatant was filtered and loaded onto 3 mL of IgG resin (Sigma-Aldrich) equilibrated in 1X lysis buffer (50 mM Tris-HCl pH 7.38, 30 mM NaCl, 1 mM EDTA, 10% glycerol, 1% DMSO, and 10 µM $ZnCl_2$) at a flow rate of about 1 mL/minute at 4 °C. After loading, the resin was washed with 50 mL of 1X lysis buffer containing 500 mM NaCl, and then equilibrated into TEV cleavage buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 10 µM $ZnCl_2$, 20% glycerol, 0.5 mM EDTA, 2 mM EGTA, 1 mM DTT, and 0.5 mM PMSF). Exosomes were cleaved from the resin by application of 100 µL of 1 mg/mL TEV protease followed by incubation of 1 hour at 4 °C. Normally, three rounds of cleavage were sufficient to completely remove the complex from the resin. Elutions were analyzed by SDS-PAGE, and then pooled and dialyzed overnight at 4 °C into 20 mM Tris pH 8.0, 20% glycerol, and 1 mM DTT. Protein was loaded onto a 1 mL MonoQ column and run in the same buffer, and eluted with an increasing gradient of NaCl. Peak fractions were analyzed by SDS-PAGE (Fig. 3-5A). Peak fractions were concentrated and flash frozen in liquid nitrogen and stored at -80 °C until use.

Preparation of Mutant Rrp44 Exosomes

To reconstitute exosomes with mutant Rrp44, wild type exosomes were prepared from yeast in a similar manner, but the resin was additionally washed with 40 mL of 1X lysis buffer containing 1.2 M $MgCl_{2.,}$ which was adjusted with 1M Tris to a pH of 7.38 to prevent a decrease in pH from the addition of magnesium, which could result in proteins leaching from the column. This wash step efficiently removed endogenous wildtype Rrp44 from the exosome complex (Fig. 3-5B). Core exosomes were then cleaved with TEV protease, dialyzed into low salt buffer, pooled and concentrated. Core exosomes were mixed with recombinant mutant Rrp44 (D551N, D171N, or D551N-D171N) in about a 1:1.5 molar ratio, incubated for 3 hours at 4 °C, and separated on a 1 mL MonoQ column in buffer containing 20 mM Tris pH 8.0, 20% glycerol, 1 mM DTT, and eluted using an increasing NaCl gradient. Excess Rrp44 eluted at about 200 mM NaCl, while exosome complexes eluted at 300 – 350 mM NaCl (Fig. 3-5C).

3.3.2. Results

Wild-type yeast purified exosome were well assembled as judged by SDS-PAGE (Fig. 3-5) and were active in biochemical assays (sections 3.4-3.6), and for EM (*19*) and single molecule (*20*) studies. Exosomes containing mutated Rrp44 lacked any wild-type activity (sections 3.4 and 3.6), indicating that no free Rrp44 was carried over from purification. Overall, the methods detailed in this section were successful for producing quality exosomes for analysis.

3.4. Biochemical assays using mutant Hepatitis Delta Virus ribozyme substrates

Initial studies used the Hepatitis Delta Virus (HDV) ribozyme. The HDV ribozyme is a self

cleaving ribozyme that catalyzes site specific cleavage in the hepatitis delta virus (*22*, *23*). Crystal structures of the wildtype (*24*) and mutant (*25*, *26*) ribozyme indicate that it forms a stable three dimensional structure, and thus for our purposes it serves as a model structured RNA substrate.

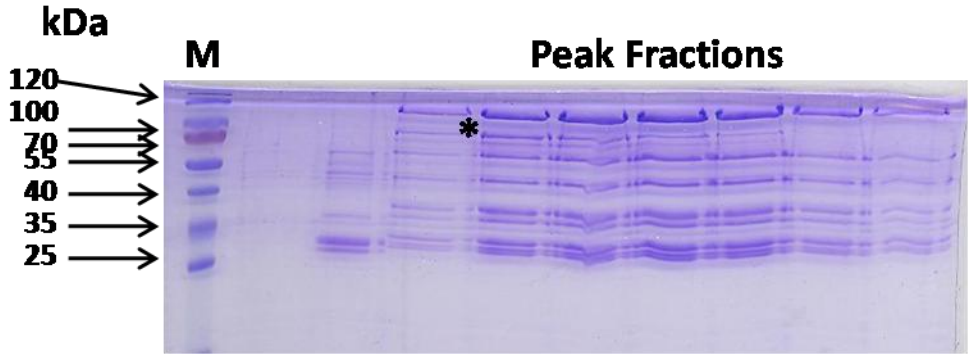
3.4.1. Methods

Hepatitis Delta Virus ribozyme RNA Preparation

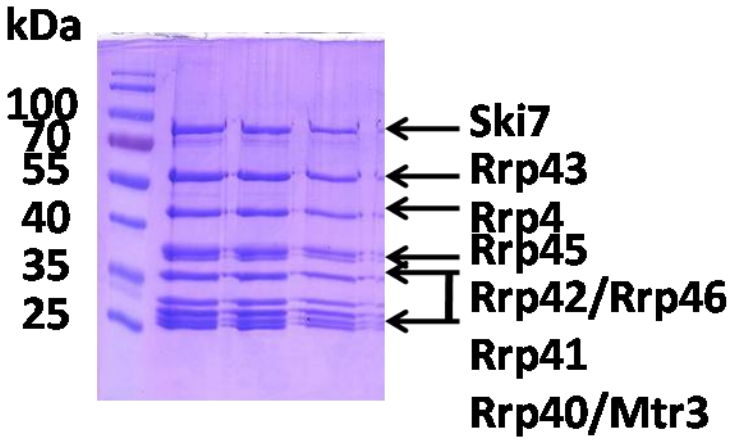
RNA substrates were engineered to contain a U1A protein binding loop and designed to contain a 3' end tail of 0 – 80 nucleotides consisting mostly of adenines, with one uridine placed between a length of ~4-5 adenines. Additionally, the template contains a C75U mutation to render the ribozyme catalytically inactive. All substrates were prepared by *in vitro* transcription using T7 RNA polymerase as follows. A 3 mL transcription reaction was preformed for each substrate using 150 µg of DNA template, 500 µL of recombinant T7 RNA polymerase diluted 1:20, 150 µL each of 100 mM ATP, UTP, GTP, and CTP; 30 mM Tris pH 8.1, 10 mM DTT, 25 mM magnesium acetate, 2 mM spermidine, 0.01% TritonX100, and diluted up to 3 mL with deionized water. The reactions were incubated for 6 hours at 37 °C, mixed with an equal volume of 2X RNA loading dye (95% formamide, 0.025% bromophenol blue, 0.025% xylene cyanol FF, and 10 mM EDTA), and purified on 10% polyacrylamide/8 M urea gels. RNA was visualized with UV shadowing. The RNA was excised from the gel, the gel piece was crushed, and RNA was eluted in nuclease free deionized water at 4 °C overnight. Substrates were dephosphorylated with FastAP Alkaline Phosphotase (Fermentas), and extracted twice with Tris saturated phenol and once with chloroform to remove all traces of enzyme. RNA was then ethanol precipitated with 50 mM NaCl, 80% ethanol, and 2 µL glycoBlue (Ambion) at 4 °C. RNA was centrifuged

Figure 3-5. SDS-PAGE analysis of yeast purified wildtype and mutant exosomes. A) TAP purification of wildtype exosomes followed by MonoQ polishing. Note that the asterisk represents Ski7, which is present in the cytoplasmic exosome and substoichiometric in all purifications. It is not a nuclease and does not affect subsequent assays. B) Removal of Rrp44 after washing with 1.2 M $MgCl_2$. C) Reconstitution of the core exosome with mutant Rrp44 after MonoQ purification. Shown is the reconstitution with Rrp44 D551N. Free Rrp44 elutes early in the run (~200 mM NaCl), while the Rrp44 bound exosome elutes later (~300 – 350 mM NaCl), indicating that exosome are not contaminated with free Rrp44.

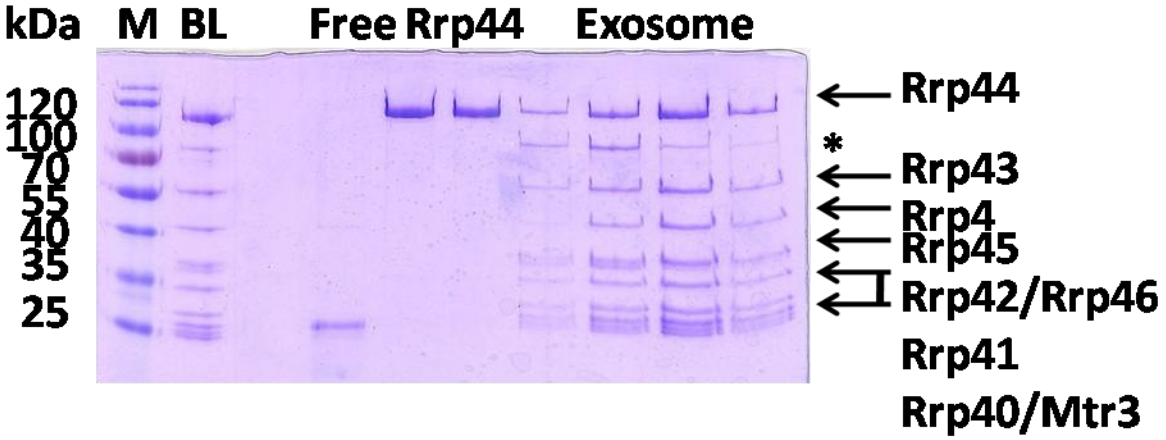
A


kDa
M
Peak Fractions
120
100
70
55
40
35
25
*

B


kDa
100
70
55
40
35
25
Ski7
Rrp43
Rrp4
Rrp45
Rrp42/Rrp46
Rrp41
Rrp40/Mtr3

C


kDa
M
BL
Free Rrp44
Exosome
120
100
70
55
40
35
25
Rrp44
*
Rrp43
Rrp4
Rrp45
Rrp42/Rrp46
Rrp41
Rrp40/Mtr3

for 30 minutes at 13,200 rpm, ethanol was decanted, and RNA pellets were dissolved in 30 µL of nuclease free water. RNA was then 5' end labeled using polynucleotide kinase and 2 µL of 3000 Ci/mmol ^{32}ɣP-ATP (Perkin Elmer), purified on 10% polyacrylamide/8 M, and eluted and ethanol precipitated in a similar manner as mentioned earlier. RNA pellets were washed with 80% ethanol, air dried, and dissolved in a final volume of 30 µL nuclease free water, and stored at -20 °C until use.

RNA Refolding

Directly before conducting activity assays, RNA was refolded in 20 mM HEPES pH 7.5, 100 mM NaCl, 200 µM $MgCl_2$ by heating to 80 °C for 5 minutes followed by snap cooling at 4 °C. All refolded substrates appeared to migrate as a single species when analyzed 8% native acrylamide/2.5 mM magnesium acetate gels (not shown).

RNase Assay

For exonuclease assays, RNA was freshly refolded and diluted to reduce the magnesium concentration to 100 µM. Enzyme was added and incubated in a time course at 30 °C. For time course assays, concentrations of 100 nM enzyme and 100 nM substrate were used. For endonuclease assays using Rrp44 D551N, $MnCl_2$ was added at a concentration of 3 mM after refolding, since heating RNA in high manganese otherwise results in RNA backbone cleavage. Reactions were stopped by the addition of 2X RNA loading dye, heated for 10 minutes at 65 °C, and loaded onto 12-16% Acrylamide Urea Sequencing Gels (National Diagnostics). Radioactivity was detected using with a phosphostorage screen, and developed using a phosphoimager (Typhoon, Amersham).

Hydroxyl radical (OH) RNA ladders were creating by incubating RNA with an equal volume of buffer containing 100 mM sodium bicarbonate pH 9.0 and 2 mM EDTA at 80 °C for 5

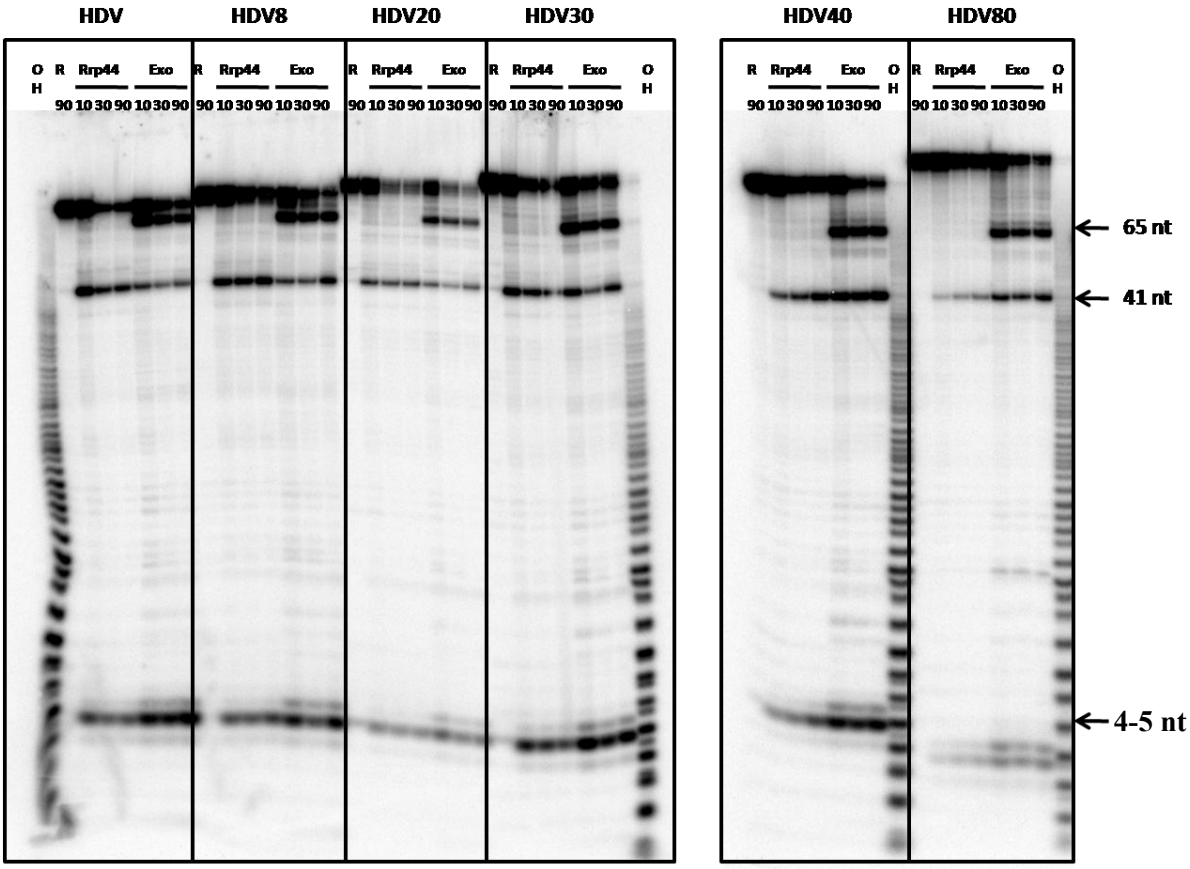
minutes. RNase T1 ladders were produced by incubating 10^{-2} units/μL of RNase T1 with RNA in buffer containing 20 mM sodium citrate pH 3.5, 1 mM EDTA, 50 mM NaCl, and 7 M urea at 50 °C for 15 minutes, followed by extracting twice with Tris saturated phenol, once with chloroform, and precipitating with ethanol.

3.4.2. Results

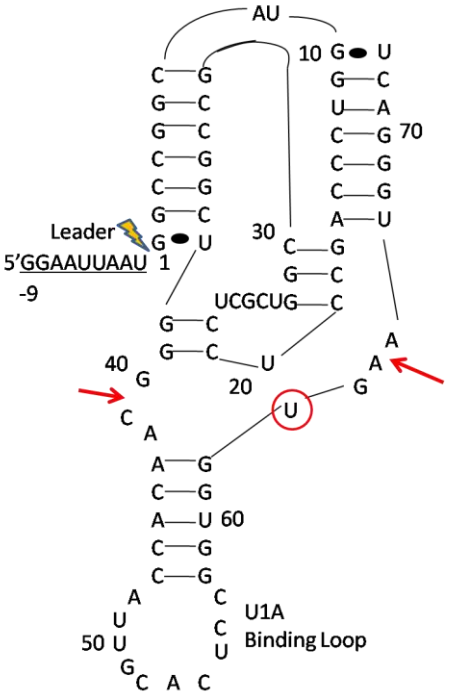
Initial assays were conducted by incubating 100 nM of Rrp44 or TAP purified exosome with 100 nM of various RNA substrates in buffer containing 100 μM $MgCl_2$ at 30 °C. Somewhat surprisingly, HDVs with 3' end tails of 0, 8, 20, 30, 40 and 80 nucleotides all produced identical intermediates and end products regardless of tail length (Fig. 3-6A). For reactions with Rrp44, an incomplete digestion occurs, producing an intermediate of ~41 nt as well as end products of 4-5 nucleotides for all HDV substrates tested. Using 5 – 10 fold more Rrp44 resulted in an almost complete digestion of the test HDV40 substrate and a larger accumulation of intermediates and end products compared to the former assay (Fig. 3-7A). Digestion of the HDVs with the Rrp44-bound exosome appears to proceed at a comparable rate and results in the accumulation of both the ~41 nucleotide intermediate as well as end products, but also produces an additional intermediate at ~65 nucleotides, which represents the size of the HDV that has had the 3' end tail removed. A titration of exosome concentration indicated that even 10 fold more exosome than substrate still resulted in an incomplete reaction (Fig.3-7A). A time course degradation of the exosome at various temperatures using HDV80 as a substrate revealed that at even 4 °C, the 3' end tail is removed within a matter of minutes, while the other intermediate and end products are produced slower (Fig. 3-7B).

Figure 3-6. Degradation Assay of HDV 3' end extensions with wildtype Rrp44 and wildtype Rrp44 bound exosome. The mutant HDV is degraded by Rrp44 or the exosome in the same manner regardless of 3' end tail length. A) Degradation assays with Rrp44 and the exosome on HDV substrates with 3' end tail lengths of 0 – 80 nucleotides. B) Cartoon representation of the secondary structure of the HDV ribozyme. Vertical dashes represent Watson-Crick base pairs and ovals represent wobble base pairs. The normal HDV cleavage site is marked with a lightning bolt, and the mutated catalytic residue is circled. Red arrows mark sites of intermediates for Rrp44 and/or exosome mediated degradation. C) Mutational studies of Rrp44 and exosome mediated degradation using HDV30 as a model substrate.

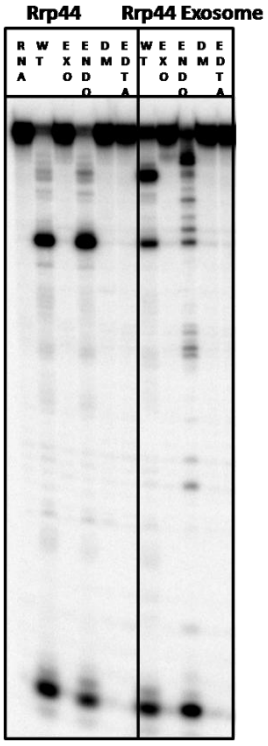
A.


HDV
HDV8
HDV20
HDV30
HDV40
HDV80
Rrp44
Exo
65 nt
41 nt
4-5 nt

B.


Leader
5'GGAAUUAAU
U1A
Binding Loop

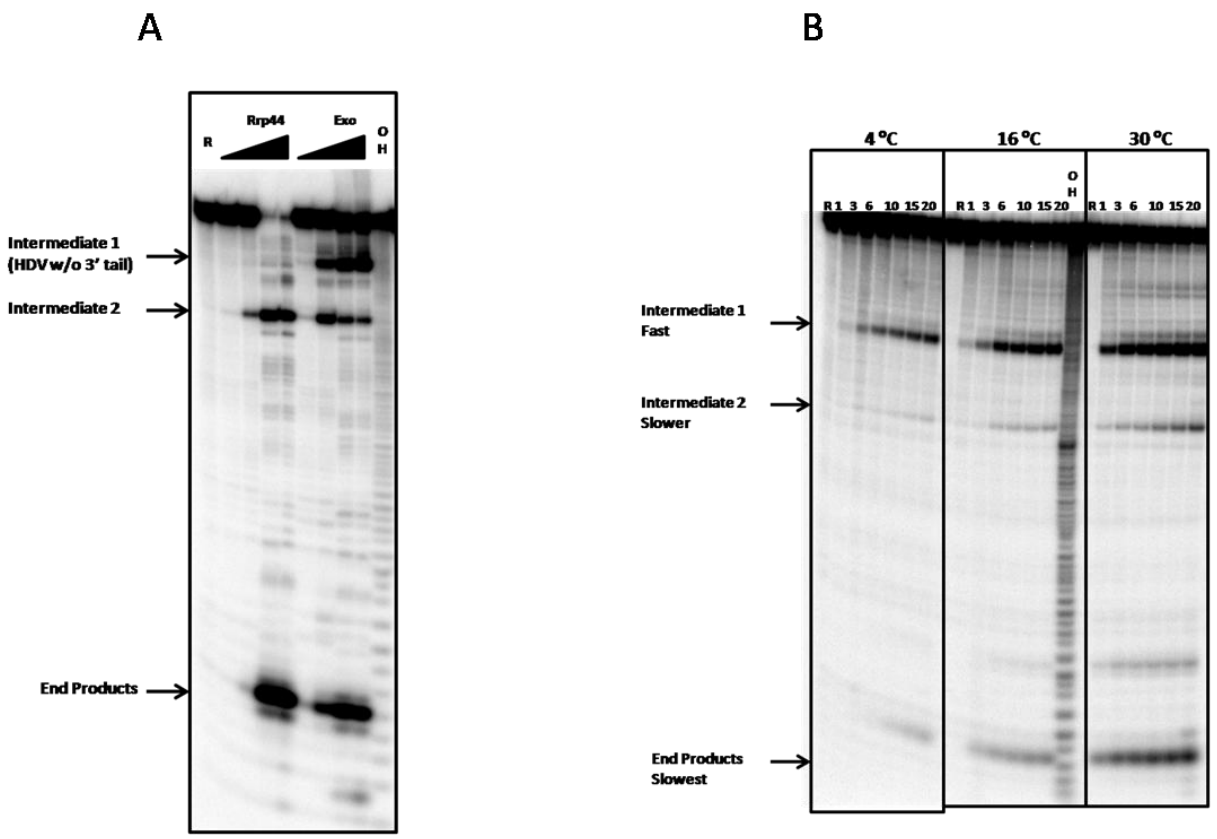
C.


Rrp44
Rrp44 Exosome

To test whether the RNase activity of Rrp44 and the exosome occurred via an exo or endonucleic pathway, mutations were made in Rrp44 in the exonuclease site (D551N, ex), endonuclease site (D171N, en), or both sites (D551N-D171N, dm) and HDV40 was assayed under the same conditions as tested previously using both Rrp44 and the exosome. Both wild-type Rrp44 and the exosome, as well as the endo mutants, degraded the RNA predominantly by the exonucleic pathway under Mg^{2+} conditions, while the exo and double mutants had no activity (Fig. 3-6C). Thus, the results of this study indicate that Rrp44 alone has the ability to unwind and degrade even highly structured RNA via its exonuclease activity under concentrations of low magnesium. The Rrp44 bound exosome displays a similar cleavage pattern, yet degradation is impeded and never goes to completion. This indicates that the bound exosome core partially prevents Rrp44 from degrading through the entire substrate.

To investigate how Rrp44 and the exosome deal with "roadblocks" such as bound proteins during degradation, a degradation assay was conducted in the presence of a titration of the U1A protein, which binds to an engineered site on the HDV substrate (Fig. 3-6B). HDV was incubated with increasing amounts of U1A for 30 minutes on ice before the addition of either Rrp44 or the Rrp44 bound exosome, and then incubated at 30 °C for 90 minutes. U1A binding effectively impedes degradation, most notably for the exosome, which had a dramatic reduction of both intermediates and end products under high U1A concentrations (Fig. 3-8A). Interestingly, high concentrations of U1A also appear to halt Rrp44 from degrading into the interior of HDV, and produce a first intermediate of ~65 nt that runs at the same relative position as the intermediate originally produced only during exosome mediated degradation (Fig. 3-8A).

Figure 3-7. Degradation of mutant HDV substrates at various temperatures and protein concentrations. A) Degradation assay of HDV40 using a titration of 10 nM – 1 μM of either Rrp44 or exosome. B) Degradation assay of HDV80 performed at the various temperatures indicated.

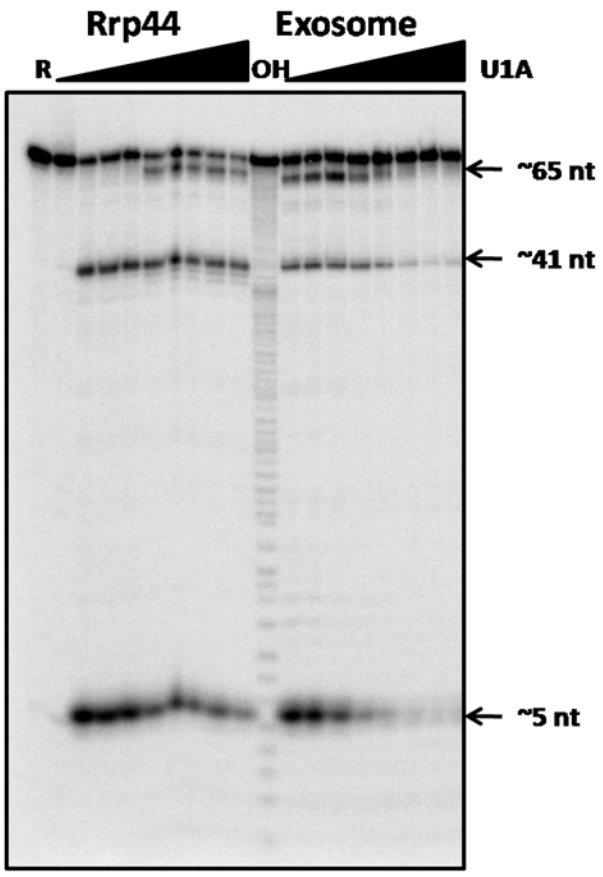

A
Rrp44
Exo
R
OH
Intermediate 1
(HDV w/o 3' tail)
Intermediate 2
End Products
B
4 °C
16 °C
30 °C
R 1 3 6 10 15 20
OH
Intermediate 1
Fast
Intermediate 2
Slower
End Products
Slowest

Since the U1A binding loop on HDV lies between the sites of the two intermediates (Fig. 3-6B, nucleotides 48 – 57), it appears that steric hindrance upon U1A binding partially impedes Rrp44 from degrading the HDV past the U1A binding site. The more dramatic reduction in activity witnessed when using the exosome indicates a possible regulatory role of the core exosome in allowing for the degradation of accessible 3' ends of substrates, yet preventing degradation into regions of RNA bound by proteins. It is therefore likely that physiological exosome substrates that form ribonucleoprotein particles like snoRNAs are prevented from being completely degraded during maturation by the exosome by the assistance of bound proteins that block the exosome from degrading past a certain region.

It has previously been reported that the Rrp44 bound exosome is optimally active at concentrations of ~100 μM $MgCl_2$, yet inhibited at concentrations of 5 mM $MgCl_2$, which is a bit odd considering than many other nucleases are active in these conditions including Rrp6 (*16*) and that many structured RNA would not be completely folded in these concentrations. To investigate the effect of magnesium concentration on the activity of the exosome, we conducted degradation assays with HDV40 using a titration of magnesium. Similar to reports by other groups with different substrates (*27*), degradation activity of Rrp44 was only slightly reduced at 5 mM $MgCl_2$ for HDV40, while the exosome was inhibited to a greater extent (Fig. 3-8B). Since the HDV requires divalent cations for proper folding, it was possible that a lower concentration of magnesium resulted in improper folding and thus a greater likelihood that the exosome could degrade the substrate. Therefore, a more detailed magnesium titration degradation analysis was conducted with several HDV substrates that were refolded and then diluted two fold to produce $MgCl_2$ concentrations of 100 μM, 1 mM, and 2.5 mM. RNA was then incubated with either

Figure 3-8. Activity Studies with Rrp44 and the Exosome. A) U1A blocking assay on the HDV substrate. U1A was titrated at concentrations of 0, 0.1, 0.2, 0.4, 0.6, 0.8, 1, and 1.2 µM. The degradation assay was then run with either Rrp44 or the exosome. B) $MgCl_2$ concentration gradient. Degradation assays were run with either 0.05, 0.10, 0.50, 1, 5, or 10 µM, or in the presence of 10 mM EDTA (indicated as "E"). "R" represents RNA only lanes, and "OH" is the hydroxyl radical ladder.

A



B

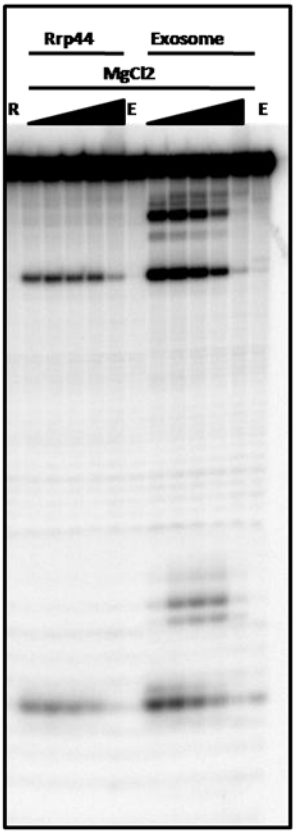
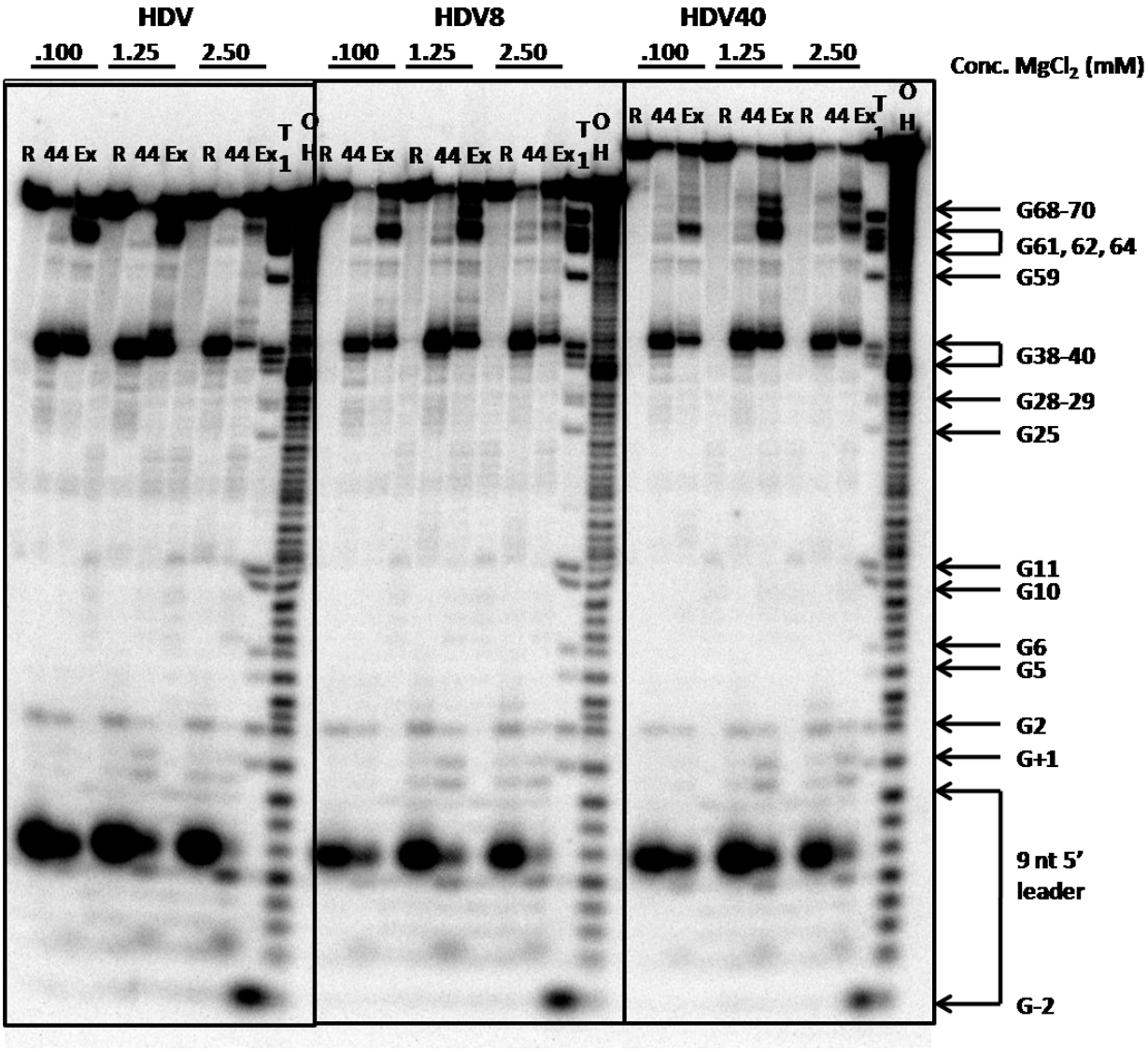


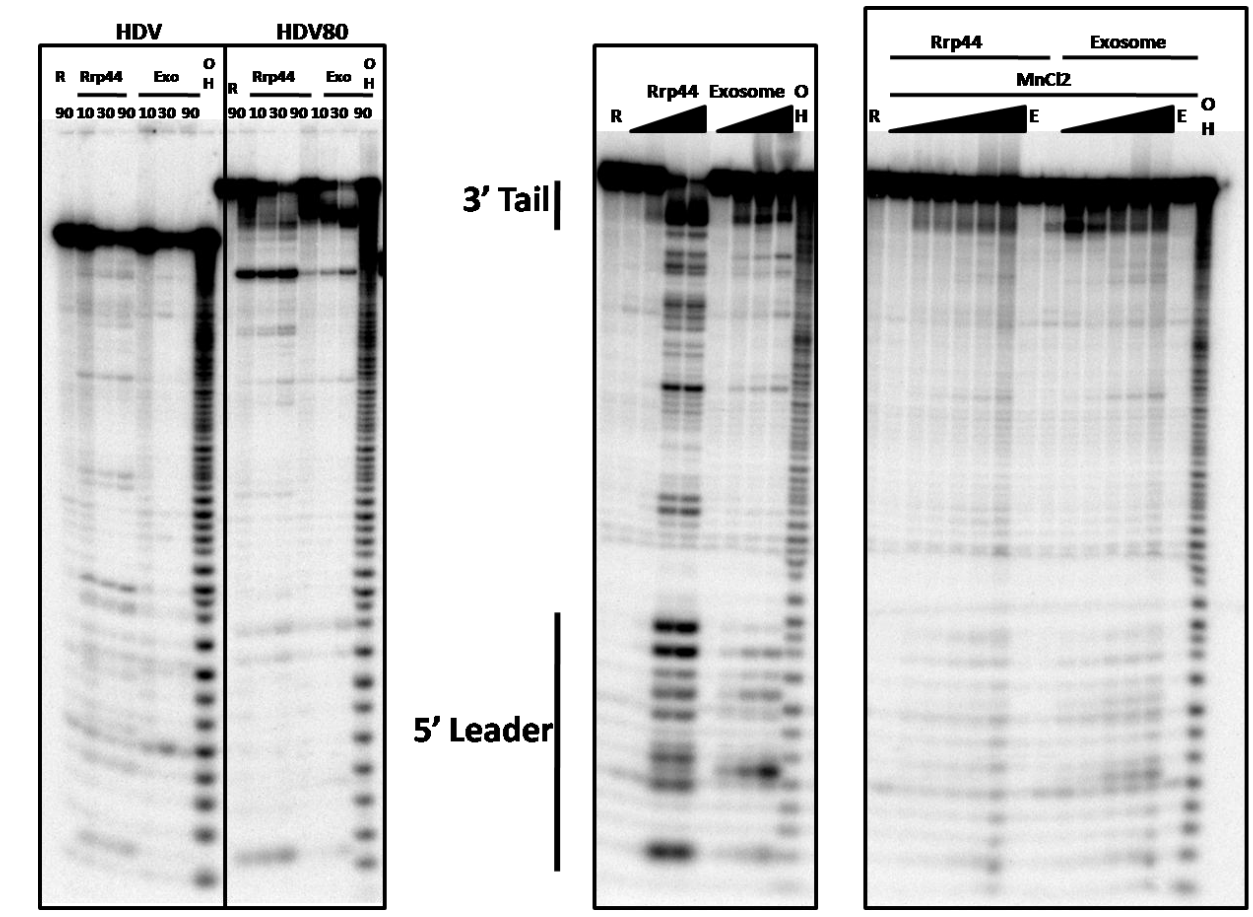
Figure 3-9. A high concentration of magnesium causes the exosome to degrade the 3' end of substrates more specifically, but not Rrp44. HDV, HDV8, and HDV40 was assayed for degradation by either Rrp44 or the exosome in the concentrations of $MgCl_2$ indicated. At high magnesium concentrations, the exosome degraded more specifically around the 3' end of the substrate, but minimal change was seen for Rrp44 alone.


HDV
HDV8
HDV40
.100
1.25
2.50
R 44 Ex
T1
OH
Conc. MgCl2 (mM)
G68-70
G61, 62, 64
G59
G38-40
G28-29
G25
G11
G10
G6
G5
G2
G+1
9 nt 5'
leader
G-2

Rrp44 or Rrp44 bound exosome in an over-digestion of 150 minutes at 30 °C. While no distinct inhibition of Rrp44 alone was displayed at these magnesium concentrations, the activity of the Rrp44 bound exosome appeared inhibited at concentrations over 1 mM (Fig. 3-9). At 2.5 mM $MgCl_2$, the HDV and HDV8 substrates were almost completely protected from degradation by the exosome while assays with HDV40 at this concentration produce a noticeable reduction of end products and accumulation of secondary intermediates near the 3' end of the substrate. For these particular substrates, the results suggest that the reduction in degradation of the structured region of the HDV could be due to better folding at higher magnesium concentrations. However, previous reports indicate that the degradation by the exosome on single stranded RNA substrates is also inhibited in high magnesium concentrations (*27*), indicating that inhibition at very high magnesium concentrations is not due to structural hindrance by the RNA alone. Therefore, it is likely that in a physiological setting the exosome is selectively active on single stranded regions at concentrations less than 5 mM but much higher than 100 μM, since at lower concentrations Rrp44 can promiscuously degrade into structured regions as well.

In order to determine the possibility of a direct access route of degradation by the exosome, it was necessary to test whether the PIN domain of Rrp44 was catalytically active while Rrp44 was bound to the exosome. The endonuclease activities of Rrp44 and the exosome were tested with the exonuclease mutant Rrp44 (D551N) by conducting assays with 3 mM $MnCl_2$ using select RNA substrates. While HDV80, which has a very long 3' end tail, was degraded by both Rrp44 and to a lesser extent by the exosome, HDV—no 3' tail—produced faint degradation products that mapped to the 5' leader sequence with Rrp44 but no degradation products with the exosome (Fig. 3-10A). This indicates that the PIN domain of Rrp44 is active when bound to the exosome, which agrees with previous results from the Conti group (*2*).

Figure 3-10. Endonuclease activity of Rrp44 D551N and the Rrp44 D551N exosome. A) Time course of HDV and HDV80 in buffer containing 3 mM $MnCl_2$. B) Protein titration on HDV40. Substrate was incubated with a concentration of 10, 100, 500, or 1000 nM of either Rrp44 D551N or the Rrp44 D551N exosome for 90 minutes. C) Manganese titration. Reactions were run a gradient of 0.5, 1, 3, 5, or 10 mM $MnCl_2$, or with 25 mM EDTA (lanes labeled with "E").


A
B
C
HDV
HDV80
R
Rrp44
Exo
O
H
90 10 30 90 10 30 90
Rrp44
Exosome
MnCl2
E
3' Tail
5' Leader

Furthermore, higher protein concentrations resulted in more pronounced endonuclease degradation by Rrp44, but little change for the Rrp44 bound exosome on HDV40 (Fig. 3-10B). Specifically, Rrp44 appeared to robustly degrade regions in both the single stranded 3' end tail and 5' end leader sequence at high enzyme concentrations, while also showing some degradation at and around single stranded regions, which agrees with previous results (*7*). Much less activity was observed for the Rrp44 exosome, which almost exclusively degraded the 3' tail of HDV40, and to a far lesser extent the 5' leader. Unlike the inhibitory effects of high levels of magnesium on the Rrp44 bound exosome (Fig. 3-8B, Fig. 3-9), both Rrp44 and the Rrp44 bound exosome displayed an almost flat activity profile in manganese concentrations ranging from 1 – 10 mM (Fig. 3-10C).

When the results of the endonuclease assays are taken together, they suggest that long single stranded regions of RNA are most susceptible to endonuclease attack by the PIN domain of free Rrp44, while short loops, such as those contained HDV, are resistant to attack, and double stranded regions are protected. Moreover, the exosome core appears to regulate the endonuclease activity of the PIN domain of Rrp44, allowing it to degrade exposed single stranded regions of the RNA, while preventing it from degrading into structured regions and internal loops.

3.5. Degradation of wildtype HDV RNA substrates by Rrp44 and the exosome

Since both Rrp44 and the exosome were shown to degrade into the structured region of the mutated HDV ribozyme, we suspected that such degradation may arise from slight flexibility of the HDV ribozyme, possibly due to the addition of the leader that was appended to the 5' end

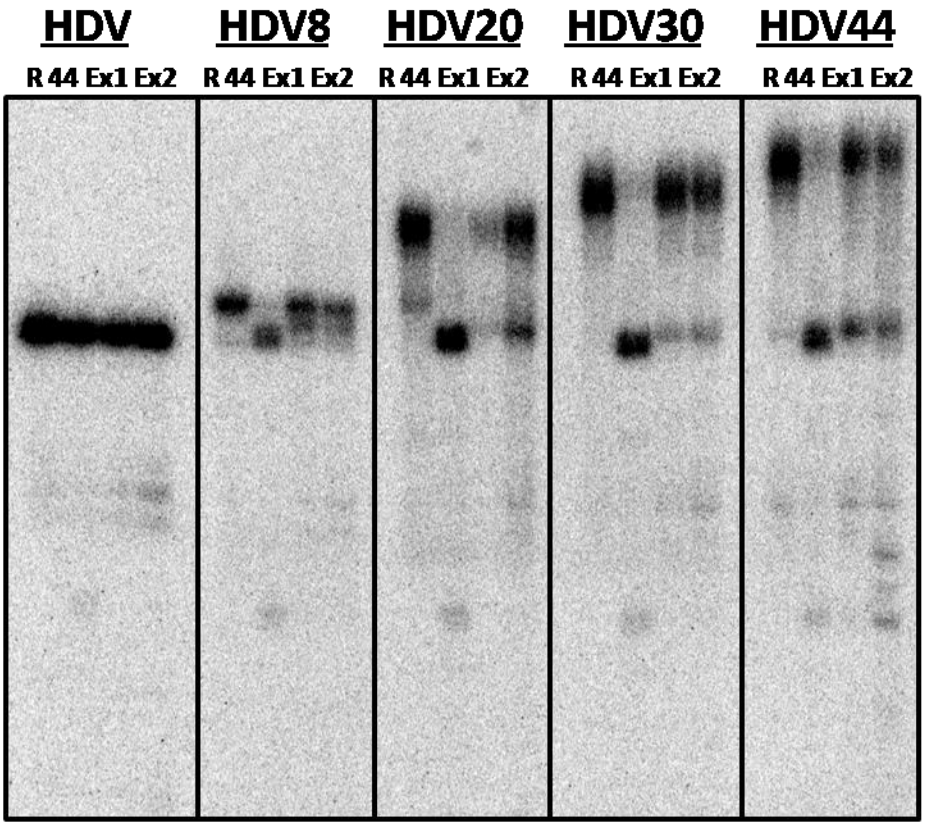


Figure 3-11. Degradation assays with wild-type HDV. Concentrations of 200 nM of RNA were incubated with either 400 nM or Rrp44 (labeled "44"), or 400 nM ("Ex1") or 1 μM ("Ex2") exosome. "R" is the RNA only lane.

of the substrates to allow for efficient radio-labeling. Therefore, to ensure that our results were accurate, we decided to test degradation using a wildtype version of the HDV ribozyme. Since the previous assays with mutant HDV ribozyme indicated that the exosome selectively degraded single stranded 3' ends at higher magnesium concentrations, a $MgCl_2$ concentration of 500 µM was used in the following assays.

3.5.1. Methods

Wildtype HDV *in-vitro* transcription

5' end labeling of the native HDV is difficult since the 5' end of the RNA is buried inside the structure when folded (*24*). Therefore, we instead body labeled the substrates during *in vitro* transcription. Small scale (500 µL) transcription was conducted in a manner similar to that described previously, but in the presence of 5 µL of 3000 Ci/mmol α-^{32}P-CTP. Transcription products were then concentrated to a volume of 40 µL using mini 10 kDa cut-off spin column concentrators. To allow for native purification, 20 µL of 60% glycerol was added to the samples before they were loaded onto a native 8% acrylamide gel containing 2.5 mM magnesium acetate and run at 600 volts for 2 hours at 4 °C. RNA was visualized by UV shadowing, and substrates were excised and eluted overnight into 20 mM Tris pH 7.5, 50 mM NaCl, and 2 mM $MgCl_2$ at 4 °C. Samples were then concentrated and stored at -20 °C until use.

RNase Assays

Assays were performed in buffer containing 10 mM HEPES pH 7.5, 50 mM NaCl with a final $MgCl_2$ concentration of 500 µM, and used a concentration of 200 nM of RNA. Rrp44 was assayed at 400 nM, and yeast exosome at concentrations of both 400 nM and 1 µM. Mixtures were incubated at 30 °C for 90 minutes and then diluted up to 100 µL with reaction buffer and

extracted with an equal volume of Tris saturated phenol. Samples were then precipitated by the addition of 450 µL of 100% ethanol and 1 µL of GlycoBlue. Samples were incubated on ice for 30 minutes, centrifuged for 30 minutes at 13,200 rpm, decanted, air dried, and dissolved in 5 µL of 2X RNA loading dye. RNA was then heated at 65 °C for 10 minutes and loaded onto a 12% acrylamide/8 M Urea sequencing gel.

3.5.2. Results

The results are what would be expected of a direct access pathway of RNA degradation by the exosome for a more structured RNA substrate. In contrast to the mutant version, the wildtype HDV lacking a 3' end extension is not degraded by either Rrp44 or the exosome (Fig. 3-11). Likewise, the 3' end tail of the HDV8 can be removed by Rrp44 but not by the exosome. However, 3' end extensions of 20 nucleotides or longer can be removed by either Rrp44 or by the exosome. As noticed earlier with the mutant HDV, exosome mediated degradation is less efficient than degradation by Rrp44 and never goes to completion. There is also a lack of any other stable intermediates or end products in this reaction, indicating the wildtype HDV appears more stable towards exonuclease attack than the mutant version.

3.6. Degradation of hypomodified tRNA by Rrp44 and the exosome

Since the degradation pattern of the wildtype HDV differed from that of the mutant HDV, a degradation study was conducted using yeast initiator methinone tRNA from yeast, which has been indicated as a physiological substrate of the Rrp44 bound exosome (*13*). Physiological eukaryotic tRNAs contain numerous posttranscriptional modifications such as the addition of a

non-template 3’ end CCA tail, as well as methylation of guanine and adenine bases and pseudouridination. The tRNAs used in this study are produced by *in vitro* transcription and have an appended CCA tail, but lack all other posttranscriptional modifications. Substrates were designed to include an additional, poly-adenine 3’ end tail of various lengths between 0 – 37 nucleotides.

3.6.1. Methods

Cloning, DNA isolation, and plasmid linearization

Initiator methionine tRNA substrates were created with 3’ end extensions of 0, 6, 10, 20, 30, and 37 adenines. The tRNA inserts were PCR amplified using overlapping primers, with a 5’ primer containing an EcoRI site followed by a phage T7 transcription start site. Inserts were then cleaved with EcoRI and inserted into a modified pUC19 vector that was linearized with EcoRI and SmaI, and then treated with polynucleotide kinase, followed by T4 DNA ligase treatment. The ligation mix was then transformed into DH5α cells, and grown on ampicillin selective LB plates. The pUC19 vector contains the wildtype HDV sequence at the 3’ flanking end of the insert, and is used in order to produce a homogenous 3’ end on the transcript, since T7 RNA polymerase has the tendency to add a few non-templated nucleotides on the 3’ end of the transcript (*28*).

Initially, DNA was isolated using miniprep and sequenced to confirm the presence of the correct insert. Correct inserts were then subsequently grown in larger volume cultures for either midi-prep or mega-prep DNA isolation, following the manufacturer’s instructions (Qiagen). Plasmids were then linearized with either HindIII or BamHI overnight at 37 °C and ethanol precipitated.

In vitro Transcription and radiolabeling

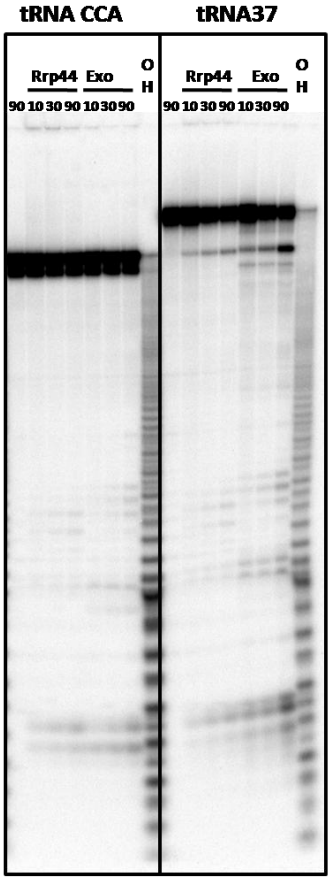
Transcription was completed as described previously (section 3.4.1) using 10 mL reaction volumes for the tRNA without 3' extension and the 6 nucleotide extension, and 1 mL volumes for the rest of the substrates. Samples were then heated at 65 °C for 20 minutes to allow the HDV to completely cleave to produce a homogeneous 3' end on the tRNA substrate. Substrates were purified by PAGE, eluted, and concentrated as described previously. Fast AP alkaline phosphate was used to remove the 5' end triphosphate as well as the 3' end cyclic phosphate created by HDV cleavage. Samples were extracted, ethanol precipitated, and 5' end radiolabeled as described previously (section 3.4.1). In general, most substrates were labeled efficiently, with the exception of tRNA10 and 20, which were poorly labeled for an unknown reason, despite the fact that all substrates were labeled following the same protocol.

RNase Assays

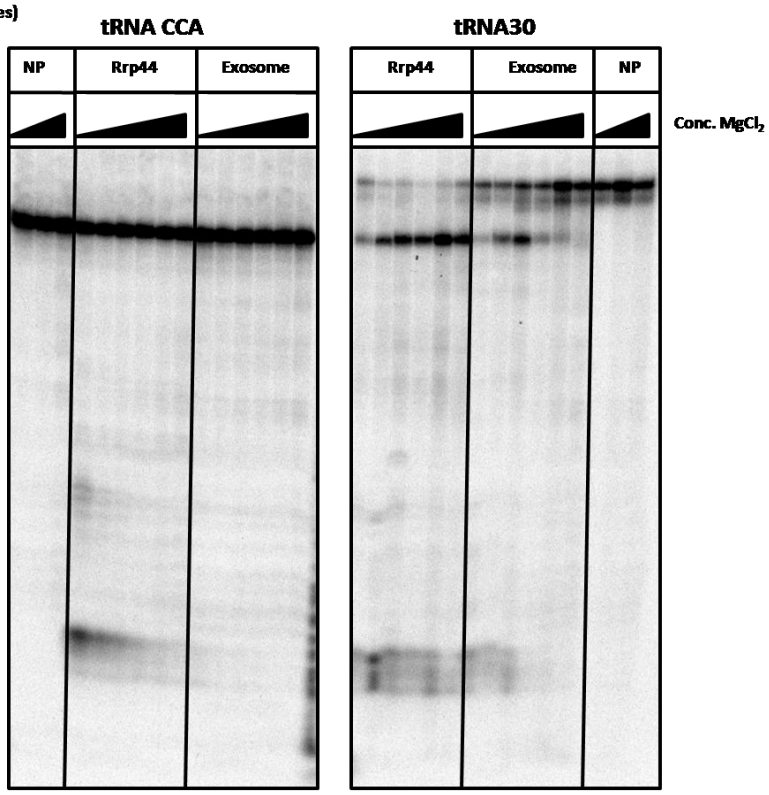
Initial exonuclease assay conditions were similar to those of the mutant HDVs. RNA was freshly refolded in buffer containing 20 mM HEPES pH 7.5, 100 mM NaCl, and between 200 µM and 20 mM $MgCl_2$ by heating to 80 °C for 5 minutes, followed by snap cooling to 4 °C. Initial assays utilized ~100 nM of RNA and 100 nM of enzyme in buffer containing 100 µM of $MgCl_2$, and degradation was tested in a time course during incubation at 30 °C. In further studies, magnesium concentrations between 100 µM - 10 mM, and protein concentrations from 200 nM – 2 µM were tested. Mutational studies were conducted with 200 nM of RNA, 400 nM of protein, and 1 mM $MgCl_2$, and incubated for 90 minutes at 30 °C. Endonuclease studies were conducted in a similar manner but with the inclusion of 3 mM $MnCl_2$. All reactions were quenched by the addition of 0.5 µL of 10% SDS, 0.5 µL of 500 mM EDTA, and 1 µL of 14 mg/mL Proteinase K, and incubated and 37 °C for 30 minutes to remove all traces of protein.

Figure 3-12. Degradation of tRNA by Rrp44 and the Rrp44 bound exosome. A) Degradation of tRNA and tRNA37 in 100 mM $MgCl_2$with Rrp44 and the exosome during a time course. Amounts of 100 nM of enzyme and 100 nM of RNA were used in this assay. B) Degradation of tRNA and tRNA30 in a magnesium chloride titration. Amounts of 200 nM substrate and 400 nM enzyme were used. $MgCl_2$ concentration ranged from 100 nM to 10 mM. Reactions were incubated for 90 minutes total. "Np" represents lanes with no protein.

A


tRNA CCA
tRNA37
Rrp44
Exo
O H
90 10 30 90 10 30 90
Rrp44
Exo
O H
90 10 30 90 10 30 90
Time (Minutes)

B

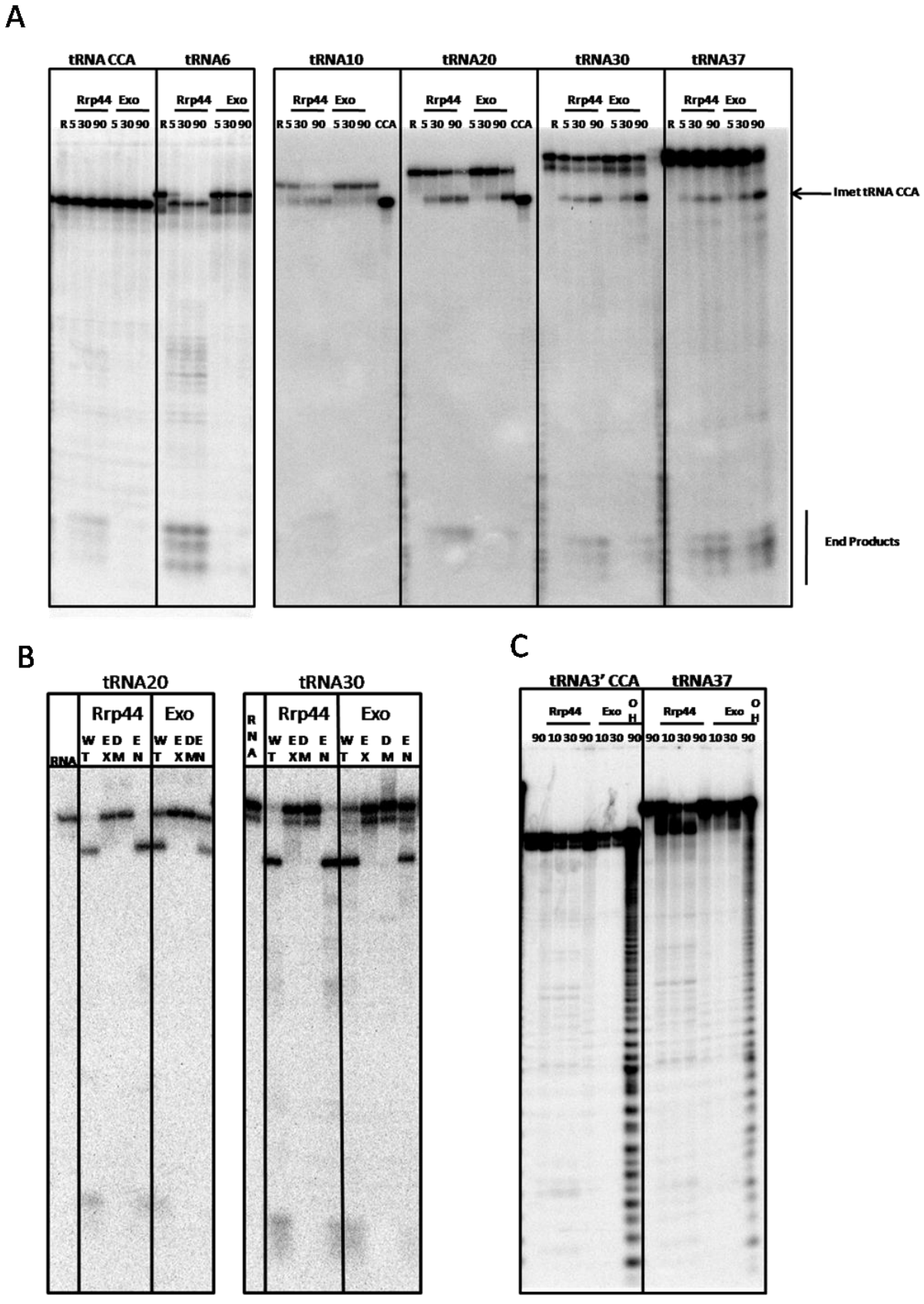

tRNA CCA
tRNA30
NP
Rrp44
Exosome
Rrp44
Exosome
NP
Conc. MgCl2

Proteinase K treatment was essential to prevent band shifts which prominently appeared for reactions containing yeast exosomes. An amount of 2-3 µL of 2X RNA loading dye was then added to the aliquots, and they were heated at 65 °C for 10 minutes. Approximately equal amounts of radioactivity were loaded onto 12% acrylamide/urea sequencing gels, and run at 1200 V, 300 mA, and 50 W for about 2 hours. Radioactivity signal was detected with a phosphostorage screen, and visualized with a Typhoon phosphoimager.

3.6.2. Results

Compared to the HDV substrate, the tRNA substrate without a 3' tail was degraded relatively poorly by both free Rrp44 and the exosome, producing only faint end products (Fig. 3-12A). However, degradation of tRNA37 with both Rrp44 and the exosome did produce an intermediate the same size of the tRNA CCA substrate, suggesting selective removal of its 3' end tail (Fig. 3-11A). To thoroughly test the effect of magnesium on degradation, a magnesium titration assay was conducted. In general, tRNA appeared more stable than wildtype HDV ribozyme at low magnesium. The tRNA without a 3' end overhang is only mildly susceptible to degradation by Rrp44 at $MgCl_2$ concentrations of 100 µM – 1 mM, producing only faint end products (Fig. 3-12B, right). The exosome showed no activity towards this substrate, regardless of the magnesium concentration. The magnesium concentration appeared less important for Rrp44 mediated degradation of the longer tRNA30, producing a main intermediate and faint end products at almost all concentrations tested (Fig. 3-12B, left). However, for exosome mediated activity, degradation down to the intermediate appeared optimal at 1 mM $MgCl_2$, and was inhibited at concentrations above this level. Overall, the exonuclease activity of Rrp44 appears relatively insensitive to the concentration of magnesium in the assay when degrading single

Figure 3-13. Degradation of tRNA substrates by Rrp44 and the Rrp44 bound exosome. A) Amounts of 200 nM of substrate and 400 nM of enzyme were used in assays containing 1 mM $MgCl_2$ and incubated during a time course. Lanes labeled as "CCA" represent lanes in which tRNA containing only a 3' CCA tail was loaded as a size marker. B) Mutational studies on tRNA substrates. C) Endonucleic tRNA cleavage by Rrp44 and the Rrp44 bound exosome. Assays were conducted with the tRNA without a 3' end or tRNA37 in buffer containing 3 mM $MnCl_2$ and incubated at 30 °C for a time course. The leftmost lane in each gel represents the RNA only sample.


A
tRNA CCA
tRNA6
tRNA10
tRNA20
tRNA30
tRNA37
Rrp44
Exo
R 5 30 90 5 30 90
R 5 30 90 5 30 90 CCA
Imet tRNA CCA
End Products
B
tRNA20
tRNA30
RNA
W T
E D XM
E N
E DE MN
C
tRNA3' CCA
tRNA37
OH
90 10 30 90 10 30 90

stranded polyadenine regions, yet is slightly impeded at higher magnesium concentrations when degrading structured regions. We find that the optimal condition for selective 3' end degradation by the exosome is about 1 mM $MgCl_2$, which would explain why previous results conducted in 5 mM $MgCl_2$ did not result in degradation (*16*). We therefore concluded that conditions of 1 mM $MgCl_2$ were sufficient to conduct further degradation assays.

Similar to titration experiments, time course assays indicate that the tRNA CCA substrate lacking a polyadenine 3' end tail was not degraded by either Rrp44 or the exosome in reaction conditions containing 1 mM $MgCl_2$. However, substrates containing 3' end tail lengths of 6 or 10 nucleotides were degraded efficiently by Rrp44 down to a product that migrated to a position consistent with that of the tRNA without a 3' end tail (Fig. 3-13A, compare to tRNA marker). Substrates with 3' end tail lengths of 20, 30, and 37 nucleotides were also degraded to the tRNA position by Rrp44. Notably, minor degradation end products were produced for some reactions—mainly those with Rrp44 alone—but they are far less intense than the stable intermediates, and may result from a small pool of RNA that remains poorly folded even after refolding, or prior degradation products that arose during labeling of the RNA. Mutational studies using catalytically inactive Rrp44 or exosome indicated that degradation resulted mainly from exonuclease activity (Fig. 3-13B).

Endonuclease assays conducted with tRNA substrates agreed well with studies with the mutant HDVs (section 3.4). Assays with tRNA and tRNA37 indicated that Rrp44 robustly cleaved the 3' end off of tRNA37, while refraining from cleaving into the interior of the substrate (Fig. 3-13C). Endonuclease activity with the exosome was only apparent in long 3' end overhang, and internal loops and structured regions appeared protected from degradation, which is consistent with results from the mutant HDV substrates. Therefore, we can

conclusively state that the exosome modulates the ability of Rrp44 to selectively endonucleotically degrade long 3' ends and avoid internal regions.

3.7. Investigating the role of the PIN domain in RNA binding.

While the PIN domain has an established role in endonuclease activity, it was also reported to be necessary for duplex unwinding of substrates with a short 3' end overhang (*3*). We investigated the role of the PIN domain in RNA binding using electrophoretic mobility shift assays (EMSA) using HDV and tRNA substrates.

3.7.1. Methods

Cloning and protein purification

Rrp44 full length and mutants were prepared as described earlier (section 3.2.1). Rrp44 242-1001 (wildtype and exonuclease mutant) and 25-1001 endo-exo mutant were cloned into pET19b vectors by using NcoI and XhoI restriction sites. These constructs were based on fragments of the protein that co-crystallized when bound to RNA (242-1001 (*3*)) or the Rrp41/45 exosome subunits (25-1001 (*2*)). Proteins were purified by Ni-NTA and MonoQ chromatography.

EMSA

EMSA was performed by incubating 200 µM of RNA (mutant HDV or tRNA) with a titration (~50 nM – 2 µM) of Rrp44 242-1001 (Rrp44ΔPIN) exo mutant or 25-1001 endo-exo mutant, in buffer containing 10 mM HEPES pH 7.5, 5 mM magnesium acetate, 10% glycerol, and 1 mM DTT for ~3 hours at 4 °C. Mixes were then resolved through 6% acrylamide gels containing 2.5 mM magnesium acetate at 4 °C, and RNA was visualized by phosphoimaging.

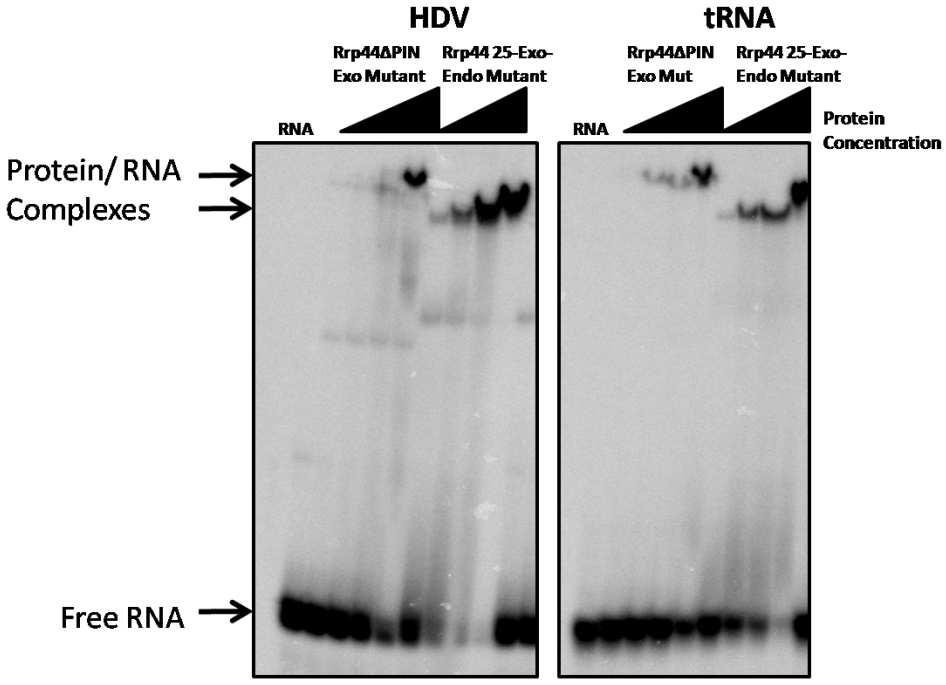
3.7.2. Results

Rrp44ΔPIN binds both HDV and tRNA weaker than Rrp44 25-1001, which contains an almost full length version of the protein (Fig. 3-14A). This, the PIN domain clearly is involved in RNA binding. For an unknown reason, the Rrp44ΔPIN/RNA complex shifts to a higher position than the Rrp44 25-1001 complex.

3.8. Discussion

The results of the biochemical assays indicate a clear inhibitory role of the core exosome on the activity of Rrp44 towards highly structured RNA substrates, as has been suggested by other groups who used more simple substrates. The association of Rrp44 with the exosome appears to impede Rrp44 from cleaving into structured regions of RNA, yet still allows the enzyme to specifically degrade long unstructured 3' ends using both exonuclease and endonuclease activities. In our studies, it is interesting to note that the exonuclease activity of the Rrp44 bound exosome was at least partially active on all mutant HDV ribozyme substrates tested, despite the presence of short or no 3' end tails on some substrates, which theoretically indicates that they should not be channeled through the exosome core based on results reported by the Conti group (*2*). Our results indicate that magnesium concentration is critical in regulating the activity of the exosome. At low magnesium concentration (~100 μM) all HDV ribozyme mutant substrates are partially degraded, suggesting that their 3' ends are at least slightly unfolded, creating a handle for the exosome to attach to and degrade the substrate using the exonuclease activity of Rrp44. Higher concentrations of magnesium restrict degradation by the exosome to mainly to the 3' end of the substrate, and are most prominent on substrates with a long overhang. A similar, though

Figure 3-14. The role of the PIN domain in RNA binding and Rrp44 activity. EMSA assay with Rrp44ΔPIN and 25-1001 catalytically inactive mutants with mutant HDV and tRNA substrates.


HDV
tRNA
Rrp44ΔPIN Exo Mutant
Rrp44 25-Exo- Endo Mutant
Rrp44ΔPIN Exo Mut
Rrp44 25-Exo- Endo Mutant
RNA
RNA
Protein Concentration
Protein/ RNA Complexes
Free RNA

less pronounced, effect occurs for tRNA substrates as well. Therefore, it is likely that the physiological magnesium concentration required for optimal exosome activity is around 0.5 – 1 mM magnesium. Lower magnesium appears to result in more promiscuous degradation by Rrp44, but magnesium concentrations up to 5 mM appears to have little inhibitory effect on the activity of Rrp44. The Rrp44 bound exosome, however, is greatly inhibited at this concentration.

We note that results from degradation assays of mutated HDV ribozymes are not entirely consistent with other substrates tested. Notably, the mutant HDV ribozyme substrates also contained a 5' single stranded leader that was included to allow for efficient labeling, while the wildtype HDV ribozyme and tRNA had recessed, double stranded 5' ends. The 5' leader could interfere with proper folding. Therefore, it is possible that the 3' ends of these substrates could be somewhat flexible, especially at low magnesium concentrations, and thus still be able to be channeled through the exosome despite the lack of a long enough 3' end tail. While assays conducted with these substrates in low magnesium result in more end products, it is quite unlikely that the substrates were unstructured enough to be channeled through the exoosme core—which can only accommodate single stranded RNA—given that prominent intermediates were produced during exonuclease activity. Also, unlike the other substrates, intermediates were different for either Rrp44 or exosome mediated degradation, suggesting a slightly different mechanism, indicating that degradation of highly structured regions was partially blocked by the exosome. Furthermore, endonuclease assays of mutant HDV ribozymes indicated that structured regions are more protected, and therefore it is not likely that these substrates are completely unfolded. Also, while the contribution of the 5' end of the substrate to degradation was initially believed to be minimal, its role has not been fully accessed. This region is a target for endonuclease attack by Rrp44, so it is possible that the 5' end could allow for an RNA binding

domain of Rrp44 to grip onto it to hold the substrate in place while exonuclease activity on the 3' end occurs.

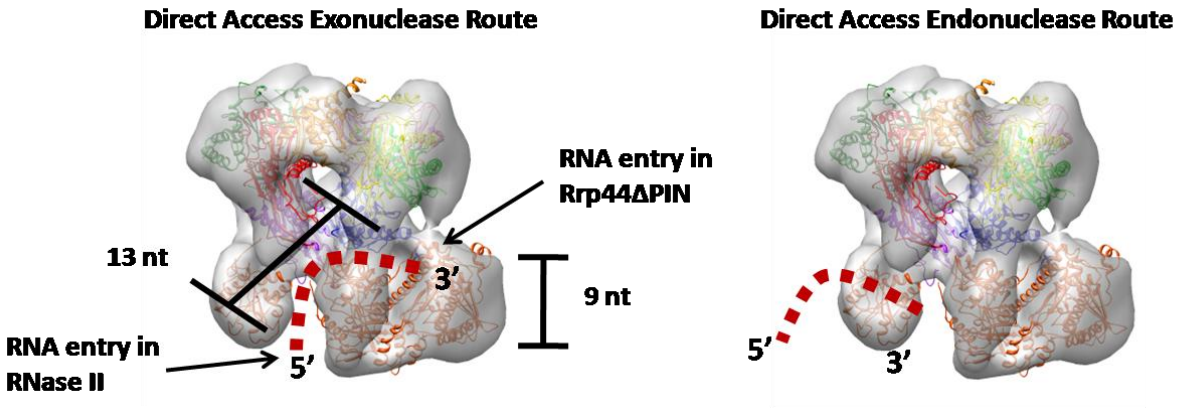
In general, consistent results from the wildtype HDV ribozyme and tRNA degradation assays point towards a pathway in which structured substrates with 3' end tail lengths of less than 20 nucleotides can be degraded by Rrp44 but not the exosome under optimal conditions. Notably, this length is 11-13 nucleotides shorter than the length of 31-33 nucleotides reported to be required to span from the opening of the exosome channel to the active site of Rrp44 (*2*). Previous protection assays indicate that the 3' end of an entirely single stranded RNA of about 20 nucleotides could travel through the exosome and reach into the active site of Rrp44, but the 5' end would be buried within the exosome channel near the exit site and it could not reach to the entry site of the exosome core (*2*). Given this evidence and the fact that wildtype HDV ribozyme and tRNA substrates contain a highly structured core that remains intact, it seems unlikely that the 3' end is being degraded by the "through exosome" route. It is more plausible that the RNA is instead being directly degraded by Rrp44, which requires only about 10-13 nucleotides to reach its active site. One caveat of this direct access pathway appears to be that it still restricts 3' end tails shorter than 20 nucleotides, indicating that the exosome impedes, yet does not completely block, direct access into the active site of Rrp44, which points towards another possible role of regulation of Rrp44 by the exosome core. The fact that a length of ~20 nucleotides on the 3' end of the substrate is required for degradation by the exosome is also consistent with a report that showed that the Trf4 component of the TRAMP complex adds about an average of 16-19 nucleotides to hypomethylated tRNA before degradation, indicating a possible physiological relevance of our results (*29*).

An inhibitory role of the exosome core on the exonuclease activity of Rrp44 on RNA has

been established previously (*27*), but our results also indicate regulation of the endonuclease activity of Rrp44 by the exosome core. While Rrp44 appears to be able to endonucleolytically degrade single stranded regions within the interior of structured RNA substrates, the Rrp44 bound exosome is mainly restricted to degrading the 3' end overhang. A likely location of the RNA substrate during degradation is hypothesized by molecular modeling. A model produced when docking the crystal structure of the ternary complex of Rrp44/Rrp41/Rrp45 and the human exosome core into the EM envelope of the yeast exosome (Fig. 2-10) suggests that the PIN domain of Rrp44 would be solvent facing when bound to the exosome and likely inaccessible to a "through exosome" RNA recruitment pathway. Therefore, the 3' end of the RNA appears to be positioned near the bottom of the exosome, which suggests that it is directly accessing Rrp44 and not being channeled through the exosome core. The mechanism of regulation of the endonuclease activity of Rrp44 by the exosome, however, is unclear because the PIN domain active site does not appear to be blocked by the exosome core. One mechanism that would be consistent with a role of the exosome in dual modulation of the exo- and endo activities of Rrp44 could be explained by the conformational change that accompanies the exonuclease region of Rrp44 upon RNA binding as displayed in the EM reconstruction (Fig. 3-1A, *right*). Such a conformational change could both restrict access of the RNA to the PIN domain and could put the RNB domain of Rrp44 in a position to access shorter 3' ends of the substrate. However, due to the low resolution of the current reconstruction, further structural studies will be required to test whether such a mechanism exists and how it occurs. Our studies also show that the PIN domain is involved in RNA binding and suggest that it can bind of RNA with RNB domain. This points towards a role of regulation of Rrp44, but further studies will need to be conducted to determine whether a similar effect occurs when Rrp44 is bound to the exosome.

When taken together, our studies point towards a direct access route of exonuclease degradation by Rrp44 when it is bound to the exosome core. The two possible routes of direct access routes to Rrp44 are the solvent exposed route based on the crystal structure of RNA bound to Rrp44ΔPIN (*3*) and the RNase II like pathway (*4*). Of these two routes, the RNase II like pathway is most supported by our results for a number of reasons. The first is that our results indicate that a 3' end overhang of 20 nucleotides is required for degradation by the Rrp44 exosome. This nucleotide length requirement is more consistent with the RNase II pathway, which requires 13 nucleotides to reach the active site, than the solvent exposed pathway, which requires only 9 nucleotides to reach the active site. While 3' end overhang lengths of 9-10 nucleotides are definitely not degraded by the exosome, it is possible that intermediate tail lengths of 13-18 nucleotides can be degraded (they have not yet been attempted) which could help to explain the current disparity between tail length requirements. However, the fact that RNA both binds to and can be degraded by the endonuclease PIN domain of Rrp44 indicates that it can in fact be recruited to the N-terminal region of the enzyme, which is in closer proximity to the entrance of the RNase II like pathway than to the exonuclease solvent exposed pathway when Rrp44 is bound to the exosome, based on EM reconstruction and subsequent crystal structure docking. Furthermore, RNA binding results in a conformational change around the RNase II like entrance pathway in the EM reconstruction. Single molecule data also implicate CSD1 as being the main RNA binding domain involved in the helicase activity of Rrp44, and this domain forms part of the entrance pathway to the RNase II like pathway, yet has no known role in the solvent exposed pathway. Future structural studies including cryo-EM reconstruction of the exosome bound to structured RNA as well as mutational studies of residues involved in the entrance site of both pathways will help to further characterize this mode of direct access Rrp44 mediated

Figure 3-15. Model of RNase II like direct access pathway to Rrp44. *Right*, The RNase II like pathway is indicated and the nucleotide length required to reach the active site (13 nt) is given, along with the nucleotide length (9 nt) required for the solvent based pathway. The data presented here more conclusively supports the RNase II pathway, which is indicated by a dashed line. *Left*, RNA can be endonucleically degraded by the PIN domain, which is close proximity to the entrance of the RNase II like direct access pathway.


Direct Access Exonuclease Route
Direct Access Endonuclease Route
RNA entry in
Rrp44ΔPIN
13 nt
3′
9 nt
RNA entry in
RNase II
5′
5′
3′

degradation by the eukaryotic exosome.

References

1. Wang, H. W., Wang, J., Ding, F., Callahan, K., Bratkowski, M. A., Butler, J. S., Nogales, E., and Ke, A. (2007) Architecture of the yeast Rrp44 exosome complex suggests routes of RNA recruitment for 3' end processing, *Proceedings of the National Academy of Sciences of the United States of America 104*, 16844-16849.
2. Bonneau, F., Basquin, J., Ebert, J., Lorentzen, E., and Conti, E. (2009) The yeast exosome functions as a macromolecular cage to channel RNA substrates for degradation, *Cell 139*, 547-559.
3. Lorentzen, E., Basquin, J., Tomecki, R., Dziembowski, A., and Conti, E. (2008) Structure of the active subunit of the yeast exosome core, Rrp44: diverse modes of substrate recruitment in the RNase II nuclease family, *Molecular cell 29*, 717-728.
4. Frazao, C., McVey, C. E., Amblar, M., Barbas, A., Vonrhein, C., Arraiano, C. M., and Carrondo, M. A. (2006) Unravelling the dynamics of RNA degradation by ribonuclease II and its RNA-bound complex, *Nature 443*, 110-114.
5. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis, *Journal of computational chemistry 25*, 1605-1612.
6. Lee, G., Bratkowski, M. A., Ding, F., Ke, A., and Ha, T. (2012) Elastic coupling between RNA degradation and unwinding by an exoribonuclease, In *Submitted*.
7. Lebreton, A., Tomecki, R., Dziembowski, A., and Seraphin, B. (2008) Endonucleolytic RNA cleavage by a eukaryotic exosome, *Nature 456*, 993-996.
8. Schaeffer, D., Tsanova, B., Barbas, A., Reis, F. P., Dastidar, E. G., Sanchez-Rotunno, M., Arraiano, C. M., and van Hoof, A. (2009) The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities, *Nature structural & molecular biology 16*, 56-62.
9. Schneider, C., Leung, E., Brown, J., and Tollervey, D. (2009) The N-terminal PIN domain of the exosome subunit Rrp44 harbors endonuclease activity and tethers Rrp44 to the yeast core exosome, *Nucleic acids research 37*, 1127-1140.
10. Liu, X., and Wang, H.W. (2011) RNA bound negative stain EM reconstruction of the yeast exosome.
11. LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A., and Tollervey, D. (2005) RNA degradation by the exosome is promoted by a nuclear polyadenylation complex, *Cell 121*, 713-724.
12. Vanacova, S., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., Langen, H., Keith, G., and Keller, W. (2005) A new yeast poly(A) polymerase complex involved in RNA quality control, *PLoS biology 3*, e189.
13. Schneider, C., Anderson, J. T., and Tollervey, D. (2007) The exosome subunit Rrp44 plays a direct role in RNA substrate recognition, *Molecular cell 27*, 324-331.
14. Callahan, K. P., and Butler, J. S. (2010) TRAMP complex enhances RNA degradation by

the nuclear exosome component Rrp6, *The Journal of biological chemistry 285*, 3540-3547.
15. Holub, P., Lalakova, J., Cerna, H., Pasulka, J., Sarazova, M., Hrazdilova, K., Arce, M. S., Hobor, F., Stefl, R., and Vanacova, S. (2012) Air2p is critical for the assembly and RNA-binding of the TRAMP complex and the KOW domain of Mtr4p is crucial for exosome activation, *Nucleic acids research*.
16. Liu, Q., Greimann, J. C., and Lima, C. D. (2006) Reconstitution, activities, and structure of the eukaryotic RNA exosome, *Cell 127*, 1223-1237.
17. Greimann, J. C., and Lima, C. D. (2008) Reconstitution of RNA exosomes from human and Saccharomyces cerevisiae cloning, expression, purification, and activity assays, *Methods in enzymology 448*, 185-210.
18. Heckman, K. L., and Pease, L. R. (2007) Gene splicing and mutagenesis by PCR-driven overlap extension, *Nature protocols 2*, 924-932.
19. Liu, X., and Wang, H. W. (2010) Preliminary Electron Microscopy Studies of Recombinant Exosomes, Yale University School of Medicine.
20. Lee, G., and Ha, T. (2010) Preliminary Single Molecules Studies of Recombinant Exosomes, University of Illinois.
21. Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Seraphin, B. (2001) The tandem affinity purification (TAP) method: a general procedure of protein complex purification, *Methods 24*, 218-229.
22. Perrotta, A. T., Shih, I., and Been, M. D. (1999) Imidazole rescue of a cytosine mutation in a self-cleaving ribozyme, *Science 286*, 123-126.
23. Perrotta, A. T., and Been, M. D. (2006) HDV ribozyme activity in monovalent cations, *Biochemistry 45*, 11357-11365.
24. Ferre-D'Amare, A. R., Zhou, K., and Doudna, J. A. (1998) Crystal structure of a hepatitis delta virus ribozyme, *Nature 395*, 567-574.
25. Ke, A., Zhou, K., Ding, F., Cate, J. H., and Doudna, J. A. (2004) A conformational switch controls hepatitis delta virus ribozyme catalysis, *Nature 429*, 201-205.
26. Ke, A., Ding, F., Batchelor, J. D., and Doudna, J. A. (2007) Structural roles of monovalent cations in the HDV ribozyme, *Structure 15*, 281-287.
27. Dziembowski, A., Lorentzen, E., Conti, E., and Seraphin, B. (2007) A single subunit, Dis3, is essentially responsible for yeast exosome core activity, *Nature structural & molecular biology 14*, 15-22.
28. Milligan, J. F., and Uhlenbeck, O. C. (1989) Synthesis of small RNAs using T7 RNA polymerase, *Methods in enzymology 180*, 51-62.
29. Kadaba, S., Wang, X., and Anderson, J. T. (2006) Nuclear RNA surveillance in Saccharomyces cerevisiae: Trf4p-dependent polyadenylation of nascent hypomethylated tRNA and an aberrant form of 5S rRNA, *RNA 12*, 508-521.

Chapter 4: Structural Studies of the Nuclear Exosome

Abstract

This chapter is focused on structural studies of the nuclear exosome feature the nuclease Rrp6. The Rrp6 containing exosome was reconstituted and analyzed by EM reconstruction. Although extra density appeared to be present for Rrp6 during 2D analysis, Rrp6 could not be located during 3D analysis, suggesting protein flexibility or dissociation. Later work involved the nuclear RNA binding protein Rrp47, and attempts were made at crystallizing the protein alone or in complex with Rrp6. While limited proteolysis was successful at producing crystals of Rrp47, no diffraction was obtained. Continued efforts will focus on engineering Rrp6 and Rrp47 in attempts at obtaining minimal regions of the proteins more readily useable for structural studies.

4.1. Introduction

The nuclear exosome is responsible for the maturation of rRNA, snRNA, and snoRNA, as well as the surveillance of mRNA and tRNA. Unlike the yeast cytoplasmic exosome, however, the nuclear exosome associates with the additional nuclease Rrp6, which appears to be involved in the degradation of the 3' precursor ends of snRNA and snoRNA, as well as some steps of rNRA maturation (*1*). The distributive nature of the exonuclease activity of Rrp6 makes the enzyme a better candidate for maturation pathways, as opposed to Rrp44, which would instead likely degrade through the substrate and could potentially unwind structured regions as well. However, Rrp6 is not an essential gene in yeast, so it is possible that Rrp44 can compensate for some of its roles *in vivo*. Current knowledge of how Rrp6 binds to the exosome core is limited to a low resolution EM reconstruction of the trypanosome exosome (*2*), and yeast two hybrid studies (*3*). Therefore, a high resolution structure of the Rrp6 bound exosome would provide much insight

into how the nuclear exosome functions.

Nuclear RNA maturation appears to involve interplay of several different proteins with Rrp6. The TRAMP complex has been proposed to activate the Rrp6 bound exosome for degradation, and the helicase Mtr4 has been shown to bind to Rrp6 in humans (*4*), but this association has not yet been established in other organisms. Moreover, two small RNA binding proteins have been implicated in playing a role in RNA maturation in concert with the nuclear exosome. Rrp47 is a 21 kDa basic protein that can heximerize and bind Rrp6 and RNA simultaneously (*5*). It is involved in the maturation of of C/D box snoRNAs, and it appears to function in part by binding to proteins in snoRNP particles, suggesting that RNA maturation occurs after particle formation (*6*). Another nuclear RNA binding protein, Mpp6, is a similar size and also highly basic. It has been implicated in maturation of 5.8S rRNA (*7*), as well as in the surveillance of pre-rRNA and pre-mRNA and in the degradation of cryptic non-coding RNA (*8*). Mpp6 has been shown to form a trimeric complex with Rrp6 and Mtr4 in human (*4*), but the conservation of such a complex and its function in the context of the exosome have not been determined.

In order to gain a better understanding of the nuclear exosome, we conducted structural studies of the Rrp6 bound exosome by negative stain EM reconstruction, and X-ray crystallography studies of Rrp47 and Rrp47-Rrp6 sub-complexes. We also conducted pull-down assays and size exclusion chromatography experiments with various reported exosome associated proteins to determine their mode of interaction. These studies were conducted in attempts of better characterizing the composition of the nuclear exosome during RNA degradation.

4.2. Purification of Rrp6 containing yeast exosomes

In order to study the Rrp6 bound exosome by EM reconstruction, we needed a practical protocol for purifying medium to large scale batches of the complex. We initially tried using a TAP-Rrp6

yeast strain, and following a purification method similar to that mentioned in Chapter 3.2. However, the TAP-Rrp6 strain yielded approximately half the amount of cell paste of the TAP-Rrp46 strain, and we were unable to obtain more than a small amount of the complex, which was not of sufficient purity for EM analysis. Therefore, we instead reconstituted yeast purified exosomes with recombinant Rrp6 in order to form the complex. This method had the added advantage that it also allowed us to engineer Rrp6 to remove flexible regions that could complicate EM analysis. Our truncated Rrp6 comprised amino acids 129-670. The N-terminal region of this truncation is based mainly on the border of the crystallized region of yeast Rrp6 (*9*), as well as a DISOPRED prediction (*10*), and data from a previous report that indicates that the C terminus of Rrp6 is required for exosome association, but the N-terminus is dispensable (*11*). In order to initially determine that Rrp6 amino acids 129-670 bound to the exosome without wasting precious purified yeast exosomes, we used recombinant exosomes. Although recombinant exosomes may not be as well assembled as their endogenous counterparts, they appear to bind Rrp6 in a similar fashion, as will be shown below.

4.2.1. Methods

Protein Expression and Purification

The coding region for the Rrp6 truncation containing amino acids 129-670 was cloned into a pLIC kanamycin vector containing an N-terminal hexahistidine tag and expressed in *E. coli* using standard procedures detailed in earlier chapters. It was then purified by Ni-NTA and MonoQ chromatography. Recombinant and yeast-purified exosomes were prepared as detailed previously (Chapter 3.2).

Association of Rrp6 129-670 with the exosome

The association of Rrp6 129-670 with the exosome was initially tested by mixing recombinant core exosomes with Csl4 and Rrp6 in a 1:1.5:1.5 molar ratio and then separating the mixture on a Superose 6 size exclusion column run in 20 mM Tris pH 8.0, 150 mM NaCl, 1 mM DTT at a flow rate of 0.3 mL/minute. Recombinant hexahistidine tagged Csl4 is added to the complex

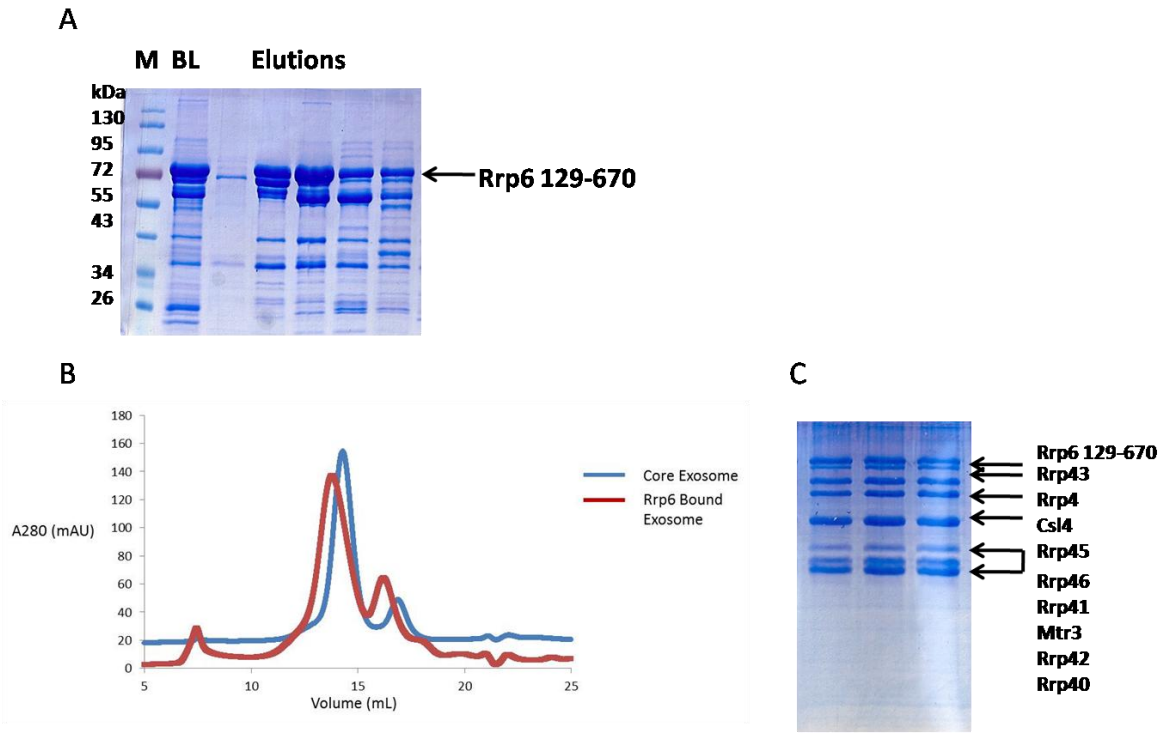
because it was absent from the previous EM reconstruction and assumed to be substoichiometric in yeast purified exosomes (*12*). The recombinant exosome + Csl4 and Rrp6 129-670 alone were run separately as size standards in a similar fashion. In order to form complexes of Rrp6 with the yeast purified exosome for EM analysis, proteins were mixed in a 2:1 molar ratio of Rrp6: exosome, loaded onto a MonoQ column equilibrated with 20 mM Tris pH 8.0, 5 mM betamercaptoethanol, and 10% glycerol, and eluted with an increasing gradient of NaCl.

4.2.2. Results

After two step purification, Rrp6 129-670 still appeared as a series of three bands that ran with close migration (Fig.4.1A), that appear to result from either protein instability or proteolysis of the C-terminus during expression. Since a smaller fragment was crystallized previously (amino acids 129-536) (*9*), it is likely that the region between amino acids 536 and 670 is unstable. However, it has been reported that this same region is important for exosome binding (*11*), so it is essential to avoid removing it for our studies. When Rrp6 129-670 was run on Superose 6 with recombinant exosomes, a peak centered at an elution volume of 13.5 mL was produced, which was an earlier elution than the core exosome alone, suggesting that an Rrp6-exosome complex was formed (Fig. 4-1B). Subsequent SDS-PAGE analysis revealed that at least one Rrp6 species was contained in the peak (Fig. 4-1C). Complex formation was also confirmed in a similar manner using the Rrp44 bound yeast purified exosome (not shown).

Ultimately, although recombinant exosomes are useful for binding studies with Rrp6, they have previously been shown to incorrectly assemble under EM (*13*). In order to produce more homogeneously assembled exosomes, we reverted to using yeast purified Rrp44 bound exosomes and reconstituting them with Rrp6 129-670. Rrp6 129-670 was mixed with Rrp44 containing yeast purified exosomes and then further purified on a MonoQ column, which was a more efficient method to simultaneously remove both RNA and unbound proteins than size exclusion chromatography. The complex eluted around 330 – 360 mM NaCl, and SDS-PAGE analysis indicated that all exosomes contained Rrp6 129-670, and approximately half of the

Figure 4-1. Purification of Rrp6 129-670 and its association with the core exosome. A) SDS-PAGE analysis of Rrp6 129-670 after Ni-NTA and MonoQ purification. About 2-3 major protein bands migrate at a position similar to the molecular weight of Rrp6, suggesting partial proteolysis during purification. "M" is the molecular weight marker, and "BL" is the before MonoQ column loading fraction. B) Elution profiles of recombinantly purified core exosome and the core exosome containing Rrp6 after size exclusion chromatography on a Superose 6 column. C) SDS-PAGE analysis of the Rrp6/core exosome complex after Superose 6 chromatography.


A
M BL
Elutions
kDa
130
95
72
55
43
34
26
Rrp6 129-670
B
A280 (mAU)
180
160
140
120
100
80
60
40
20
0
5
10
15
20
25
Volume (mL)
Core Exosome
Rrp6 Bound Exosome
C
Rrp6 129-670
Rrp43
Rrp4
Csl4
Rrp45
Rrp46
Rrp41
Mtr3
Rrp42
Rrp40

exosomes also contained Ski7 (Fig. 4-2A). Subsequent western blotting with a HISprobe antibody, which is specific for polyhistidine tags, unambiguously revealed the presence of Rrp6 129-670 in the complex (Fig. 4-2B). Interestingly, although Rrp6 129-670 initially appeared as three close migrating bands after purification, the exosome complex contains only the highest migrating Rrp6 species, which most likely contains the full 129-670 amino acids. Additionally, two species that migrate between 35 and 40 kDa are identified in the blot. One of these proteins is likely recombinant hexahistidine tagged Csl4, which migrates in this position. The other species is probably a truncated form of Rrp6 that also appears in samples of Rrp6 129-670 alone (Fig. 4-1A).

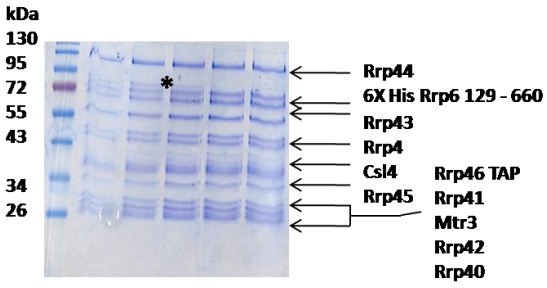
4.3. EM analysis of Rrp6 129-670 containing exosomes

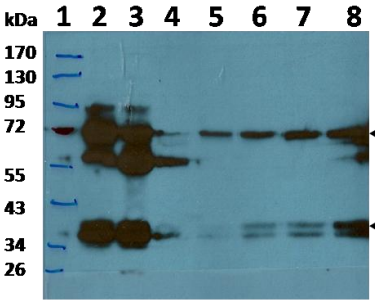
Since EM analysis requires a low concentration of sample, we had the luxury of picking single fractions from the MonoQ purification to test under EM. The advantage of this method was that it allowed us to selective avoid fractions containing the endogenous Ski7 protein, which could complicate reconstruction methods. EM analysis was conducted in collaboration with the laboratory of Hongwei Wang (Yale Medical School). In the following section, the author of this manuscript preformed initial 2D reconstruction analysis on the samples, while Xueqi Liu, a postdoctoral fellow in the lab of Hongwei Wang performed 3D reconstruction analysis.

4.3.1. Methods

Protein was diluted to ~0.1 µM in buffer containing 25 mM HEPES pH 7.5, 50 mM NaCl, and 2 mM DTT. A volume of 4 µL of protein solution was cross-linked with 0.1% glutaldehyde for either 30 seconds or 1 minute. Samples (native and cross-linked) were then negatively stained

Figure 4-2. Analysis of the Rrp6 containing exosome reconstituted from yeast proteins. A) SDS-PAGE analysis of the Rrp6 exosome after MonoQ purification. The asterisk indicates a substoichiometric amount of Ski7 that was contained in some fractions. B) Western Blot analysis of the Rrp6 exosome. Fractions from the MonoQ purification of Rrp6 129-670 (Fig. 4-1A) were compared to fractions from the Rrp6 bound exosome purification. The exosome appears to contain mostly the full version of Rrp6 (amino acids 129-670) in the peak fraction (lane 8, fraction #75). Secondary bands appear around 35 kDa: one is likely hexahistidine tagged Csl4, while the other is probably a smaller Rrp6 truncation.





1. Molecular Weight Marker
2. Rrp6 MonoQ #19
3. Rrp6 MonoQ #22
4. Rrp6 MonoQ #43
5. Yeast Exosome + Rrp6 72
6. 73
7. 74
8. 75

with 2% uranyl acetate on a copper grid, blotted dry, and processed for EM imaging. Exosome images were collected, and data processing was completed using EMAN software (*14*).

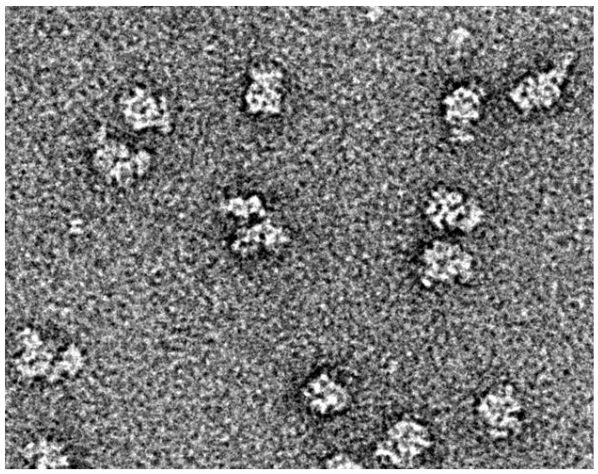
4.3.2. Results

Exosomes appeared well assembled under EM analysis (Fig.4-3). Cross-linking with a low percentage of glutaldehyde was used to allow Rrp6 to form a more stable association with the exosome and prevent flexibility. Short cross-linking allowed for a homogeneous particle size under EM (Fig. 4-3A), while cross-linking for 1 minute produced aggregation (Fig. 4-3B), and produced exosomes exosome cores that were severely constricted (not shown).

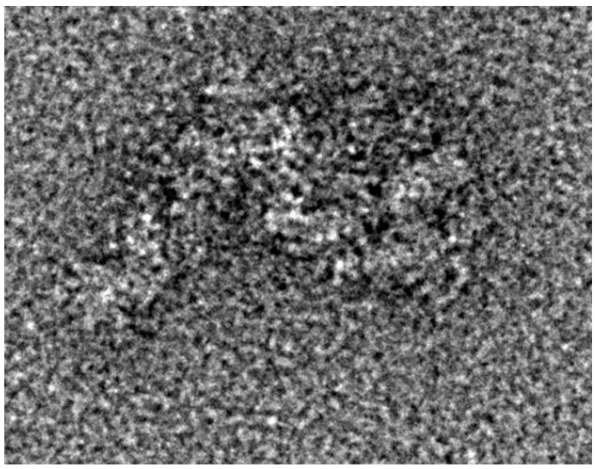
Two-dimensional class averages of the Rrp6 bound exosome were produced using EMAN software (*14*). Class averages are representative angles of view based on severalthousand individual exosome particles (Fig. 4-3C). The central channel of the exosome is very distinguishable in most of the class averages displayed. In some angles, a large extra density on the bottom of the RNase PH ring can be seen, which corresponds to Rrp44. Furthermore, quite a few views also indicate the presence of another density distal to Rrp44, which could be Rrp6 (Fig. 4-3C). Despite the promising result from the 2D class averages, three dimensional analysis was needed to conform the location and conformation of Rrp6. 3D analysis was completed by Xueqi Liu using the samples mentioned here, as well as others prepared in a similar manner under a variety of different conditions. Although an additional density was found on the top of the exosome ring in 3D reconstructions, it corresponded to Csl4 and not Rrp6, because the same density was later found in exosome preparations lacking Rrp6 (not shown). Unfortunately, the Rrp6 density was not able to be located during 3D reconstruction.

Fig.4-3. EM analysis of Rrp6 containing exosomes. No or short crosslinking produces homogeneously assembled exosomes (A), while longer crosslinking results in aggregates (B). C) Class averages obtained from Rrp6 containing exosomes. The exosome core is clear in most angles. Some angles also included a density on the bottom of the core that represent Rrp44, and a density on the top that likely was produced by Rrp6.

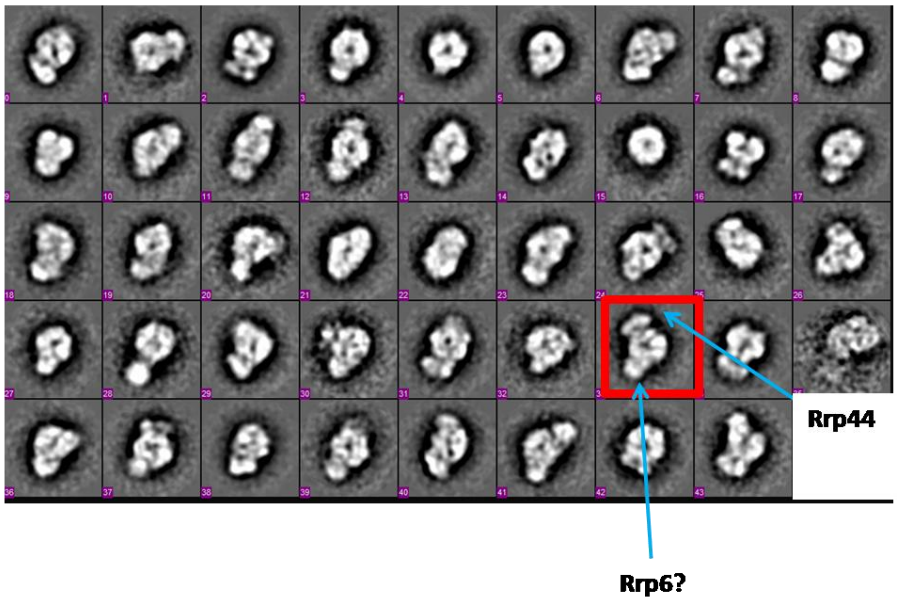
A



B



C



Western blot analysis indicates that Rrp6 is clearly present in the exosome complex (Fig. 4-2B) and 2D analysis also suggests the presence of a sizable density on the top of the exosome ring (Fig. 4-3C), which corresponds to a position where Rrp6 is likely to bind based on prior EM work (*2*) and yeast two-hybrid studies from different organisms (*3*). Therefore, the absence of Rrp6 in 3D reconstructions likely results from flexibility of the subunit, resulting in it adopting different conformations. This flexibility ultimately causes the Rrp6 density to average out during 3D reconstruction, and this is likely why we were not able to locate it.

The flexibility in Rrp6 could be reduced by a number of different methods. The first would be to further truncate Rrp6 to remove flexible regions in the C-terminus of Rrp6. This method, however, is complicated by the fact that a portion of this region appears to overlap with the binding site to the exosome (*11*). An alternative method would be to form exosomes containing additional Rrp6 binding subunits in hopes that the binding of Rrp6 to these subunits would cause it to lock into a more rigid conformation, which would improve EM analysis. In attempts towards this method, we explored the interaction of Rrp6 with other accessory nuclear proteins in the following section.

4.4. Exploring protein interactions within the nuclear exosome

Although Rrp6 is the hallmark subunit of the nuclear exosome, addition cofactors are also present. In *S. cerevisiae,* Rrp6 forms a complex with Rrp47 that is mediated by interaction between the N-terminal regions of the two proteins (*6*). Several complexes of Rrp47 and Rrp6 were produced by co-expression in an attempt to lock down flexible regions of the proteins for facilitating reconstruction efforts of the Rrp6 bound exosome. Additionally, interactions between Rrp6 and the proteins Mtr4 and Mpp6 have been reported in human cells (*4*).

Interactions between these proteins with Rrp6 and the core exosome were explored using GST pull-down assays and size exclusion chromatography analysis. Any stable interactions between these proteins were then pursued by reconstitution of the protein into yeast purified exosomes for EM analysis.

4.4.1. Methods

Cloning, protein expression, and purification

The coding sequences of the full length version of Rrp6 (amino acids 1-733) and C-terminal truncation of amino acids 1-660 were PCR amplified from yeast genomic DNA with a forward primer containing a 5' end BamHI site and an N-terminal hexahistidine encoding region, and a reverse primer with a 3' end XhoI site. Full length Csl4 was amplified in a similar manner. Rrp6 and Csl4 sequences were then inserted into a pGEX4T1 vector (containing an N-terminal GST moiety and ampicillin resistance) with T4 DNA ligase. Hexahistidine tagged Rrp6 1-660 was additionally amplified with 5' NcoI and 3' XhoI primers and inserted into a pET19b vector containing ampicillin resistant. Exosome core subunits or dimers of subunits were cloned and expressed as stated in section 3.1. Mtr4 that was cloned into a pET 151-D-topo vector that contained an N-terminal TEV protease cleavable hexahistidine tag was kindly provided by Sean Johnson (Utah State University). Coding regions for Mpp6 and Rrp47 were amplified from genomic DNA and cloned into a pSUMO vector using a 5' BamHI and a 3' XhoI restriction site. Ligation mixtures were transformed into DH5α cells. All plasmids were selected on ampicillin resistant LB plates, with the exception of Mpp6 and Rrp47 which were selected for on kanamycin plates. Plasmids containing Mtr4, Csl4, and Mpp6 were then transformed into BL21* cells for expression. GST Rrp6 full length and 1-660 (GST and native version) were co-

transformed with Rrp47 vectors for co-expression into Rosetta II cells (containing rare codons and chloremphenicol resistance) and selected on triple ampicillin, kanamycin, and chloremphenicol plates. All cultures were grown in LB supplemented with the appropriate antibiotics at 37 °C and 200 rpm until reaching an OD600 of 1.0. Cultures were then induced with 1 mM IPTG at 18 °C and 200 rpm overnight.

GST Rrp6 full length and 1-660 truncation (GST tagged and native) in complex with hexahistidine SUMO Rrp47 were purified by Ni-NTA, and dialyzed overnight into 20 mM Tris pH 8.0, 300 mM NaCl, 5 mM betamercaptoethanol, and 10% glycerol. During dialysis, the N-terminal hexahistidine SUMO tag on Rrp47 was cleaved with Sm3t SUMO protease. The complex was then further purified by heparin chromatography. Rrp47 alone was purified in a similar manner. Mtr4 and Mpp6 were purified by Ni-NTA using standard procedures, and the hexahistinde SUMO tag on Mpp6 was cleaved during overnight dialysis. The proteins were then purified on either MonoQ (Mtr4) or heparin (Mpp6) chromatography, followed by size exclusion chromatography on a Superose 6 column. For the Csl4 GST fusion, cell pellets were resuspended in 1XPBS, lysed by sonication, and applied to a 5 mL GSTrap column (GE Healthcare). The column was then washed with 20 column volumes of 1XPBS and eluted with buffer containing 50 mM Tris pH 8.0, 100 mM NaCl, and 20 mM reduced glutothionine. An empty pGEX41 vector was expressed and purified in a similar manner to serve as a control. Proteins were concentrated and flash frozen and stored at -80 °C.

GST Pulldown Assays

An amount of 50 µg of GST tagged protein (Rrp6, Csl4, or free GST as a control) was mixed with 50 µg of bait protein and diluted up to 100 µL with Binding Buffer (1XPBS containing 1 mM DTT). The mixture was then applied to 50 µL of glutathione beads in a 1.5 mL centrifuge

tube equilibrated with Binding Buffer. The mixture was incubated with the beads at 4 °C for 2 – 4 hours with gently rocking. The tube was then spun and the supernatant was collected. The beads were washed three times each with 1 mL of Binding Buffer, and eluted with 125 µL of 50 mM Tris pH 8.0, and 20 mM reduced glutathione. Elutions were run on 10% SDS PAGE gels and visualized with Coomassie Blue staining.

Size Exclusion Chromatography

Mtr4 and Mpp6 were run either separately or in combination through a Superose 6 column equilibrated with 20 mM Tris pH 8.0, 150 mM NaCl, 1 mM DTT at a flow rate of 0.3 mL/minute. Fraction sizes of 0.250 mL were collected for analysis.

Complex Formation with yeast purified exosome

The untagged Rrp6 1-660/Rrp47 complex was mixed with yeast purified exosomes in a 2:1 molar ratio and dialyzed overnight into 20 mM HEPES pH 7.5, 75 mM NaCl, and 2 mM DTT. The mixture was then applied to a 1 mL MonoQ column, run using dialysis buffer, and eluted with an increasing gradient of NaCl. Peak fractions were analyzed by SDS-PAGE to ascertain complex formation. An exosome complex with Mpp6 was attempted in a similar manner.

4.4.2. Results

Stable complexes between all Rrp6 truncations tested with full length Rrp47 were clearly formed since both proteins were present during SDS-PAGE analysis after all steps of purification, despite the fact that only Rrp47 contained a hexahistidine tag and that purification was performed in high salt conditions (300 mM NaCl) during Ni-NTA chromatography. Notably, Rrp47 aided in the purification of full length Rrp6. Full length Rrp6 alone expresses poorly

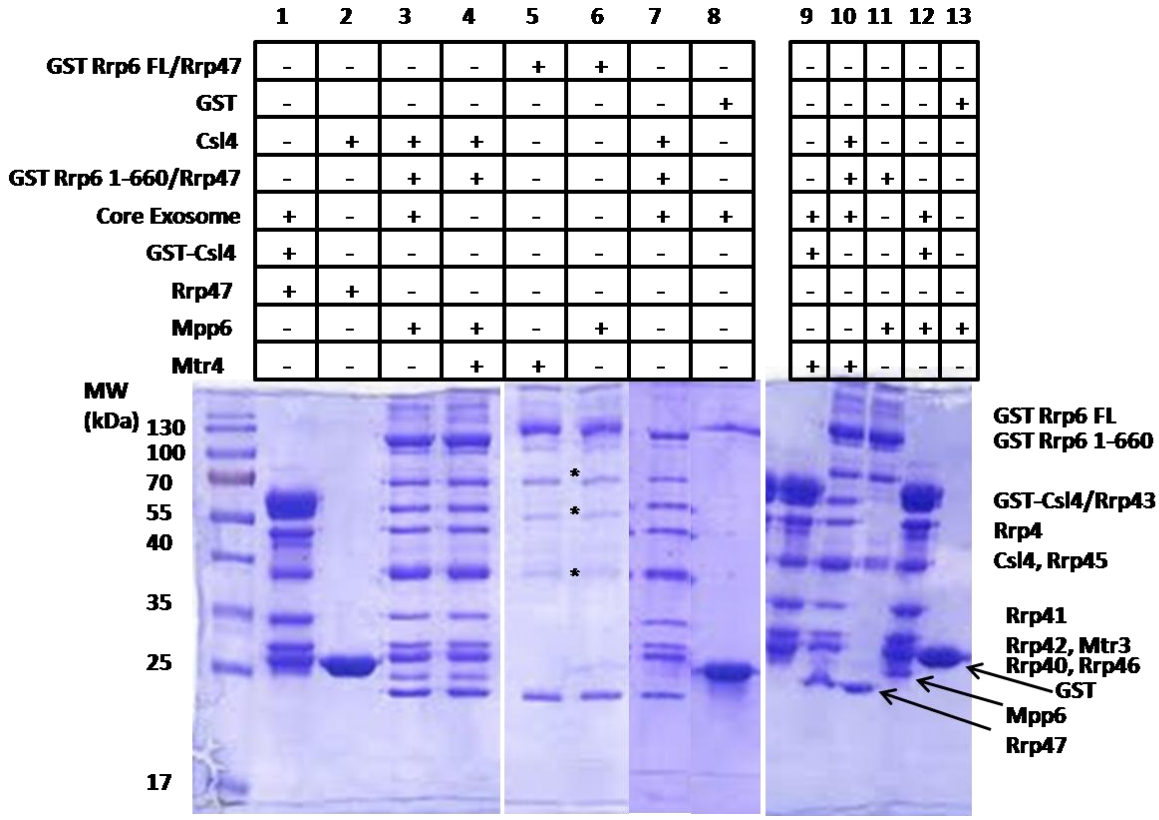
except when fused to a SUMO tag, and even then yield of the protein is very low (*15*). However, the Rrp6 full length/Rrp47 expressed well with a good yield and decent final purity.

Pull-down assays were first conducted with GST-Csl4 using the recombinant expressed exosome and Rrp47 as bait. GST-Csl4 was able to pull-down all subunits of the exosome, while GST alone did not (Fig.4-4A lanes 1 and 2), but it did not pull-down Rrp47, indicating that this protein solely interacts with Rrp6. The GST-Rrp6 1-660/Rrp47 complex was also able to pull down all exosome subunits (Fig. 4-4A lane 7), indicating that the complex likely binds to the exosome in a manner similar to the previously truncation Rrp6 129-670. The GST-Rrp6 1-660/Rrp47 complex, however, failed to pull down any of the RNA binding subunits of the exosome alone (Rrp4, Rrp40, or Csl4), or any individual heterodimer (Rrp41/Rrp45, Rrp43/Rrp46, or Rrp42/Mtr3) or Rrp44 (results not shown). Therefore, Rrp6 likely binds at the interface between different heterodimers and/or capping subunits, making the exact location of its binding difficult to decipher.

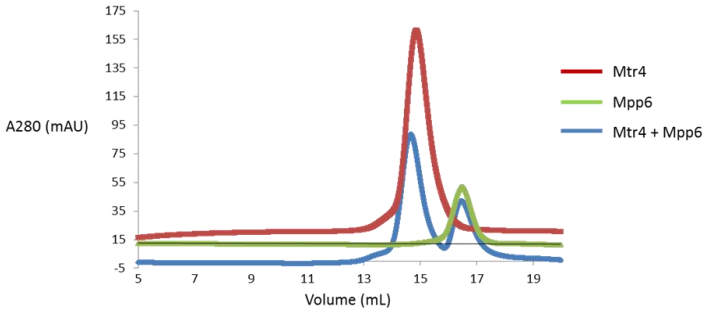
Neither the GST Rrp6 full length/Rrp47 or the GST Rrp6 1-660/Rrp47 complexes pulled down Mpp6 (Fig. 4-4A lanes 6 and 11) or Mtr4 (Fig. 4-4A lane 5) alone. The core exosome was also not able to pull down Mtr4 (Fig. 4-4A lane 9). However, GST-Csl4 pulled down Mpp6 when mixed with the core exosome (Fig. 4-4A lane 12), and a similar result was observed when GST-Rrp6 1-660/Rrp47 was mixed with the core exosome and Mpp6 (lane 3). This indicates that Mpp6 makes most of its interactions with the core exosome, and not Rrp6 in *S. cerevisiae*. There were no interactions detected between Mpp6 and Mtr4 as determined by size exclusion chromatography experiments (Fig. 4-4B-C).

Fig. 4-4. Nuclear exosome protein interaction studies. A) GST pull-down experiments, depicting fractions eluted from GST binding resin. B) Size exclusion chromatography profiles of Mtr4 and Mpp6 run on a Superose 6 column. C) SDS-PAGE analysis of size exclusion chromatography preformed by combining Mtr4 and Mpp6. No indication of complex formation was detected. The first lane of the gel contains the before loading (BL) sample.

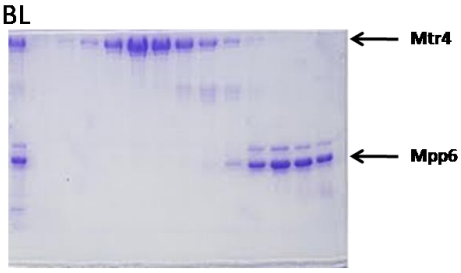
A


1 2 3 4 5 6 7 8 9 10 11 12 13
GST Rrp6 FL/Rrp47
GST
Csl4
GST Rrp6 1-660/Rrp47
Core Exosome
GST-Csl4
Rrp47
Mpp6
Mtr4
MW (kDa)
130
100
70
55
40
35
25
17
GST Rrp6 FL
GST Rrp6 1-660
GST-Csl4/Rrp43
Rrp4
Csl4, Rrp45
Rrp41
Rrp42, Mtr3
Rrp40, Rrp46
GST
Mpp6
Rrp47

B


A280 (mAU)
Volume (mL)
Mtr4
Mpp6
Mtr4 + Mpp6

C

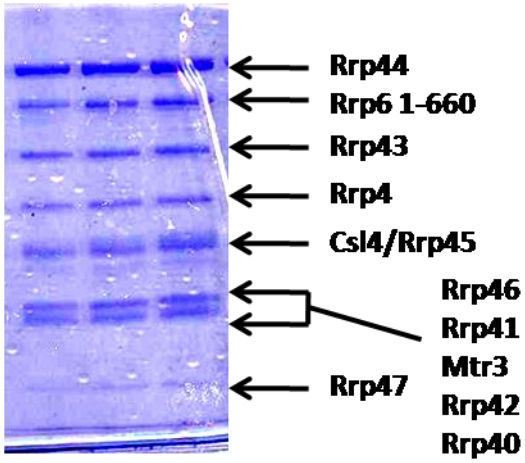

BL
Mtr4
Mpp6

Based on the results of GST pull-down and size exclusion chromatography experiments, reconstitution of yeast exosomes containing either the Rrp6 1-660/Rrp47 complex or Mpp6 was attempted. Although Rrp6 1-660 is clearly present in the exosome, the Rrp47 band appears rather faint in SDS-PAGE analysis (Fig. 4-5) so it is difficult to tell whether the protein is stoichiometric in the complex or not. It was even more difficult to indicate the presence of Mpp6 within the exosome since lower concentrations of proteins were used for this experiment (result not shown). Both complexes were subjected to EM analysis, but interpretable density was not observed for Rrp6, Rrp47, or Mpp6 (*13*). It is likely that Mpp6 is too small to produce pronounced density in EM analysis. The protein appears to run in a position suggesting a monomer instead of forming an oligomer (Fig. 4-4B, compare the 20 kDa Mpp6 to the 122 kDa protein Mtr4). It is surprising that no density was observed for Rrp47, which has been reported to form a hexamer (*5*), and thus should produce a large observable density. It is possible that Rrp47 dissociates during negative stain conditions, or is substoichiometric during intial purification, which would explain its absence during EM analysis

4.5. Structural Studies of Rrp47 and the Rrp6/Rrp47 complex

Structural studies were initiated on Rrp47, an RNA binding protein that binds to Rrp6 and is involved in processing and snRNA and snoRNA (*16*). Rrp47 contains a Sas/C1D domain that is predicted to be predominately alpha helical in nature, which is involved in nucleic acid binding (*17*). Other than its role in exosome associated degradation, Rrp47 has been implicated in regulation of DNA repair in mammals (*18*). Rrp47 is of particular interest because the structure of the Sas/C1D domain has not been determined from any protein homologue.

Figure 4-5. Reconstitution of the Rrp6 and Rrp47 containing exosome. Yeast exosomes were mixed with recombinant Rrp6 1-660/Rrp47 complex and purified on MonoQ chromatography. Peak fractions were analyzed on SDS-PAGE. While Rrp6 1-660 is easy to identify, Rrp47 is barely noticeable in the gel.


Rrp44
Rrp6 1-660
Rrp43
Rrp4
Csl4/Rrp45
Rrp46
Rrp41
Mtr3
Rrp42
Rrp40
Rrp47

Although Rrp47 has a molecular weight of only 21 kDa, it migrates at a position closer to 120 kDa during size exclusion chromatography (*5*), indicating that it likely forms a hexamer. Our primary structural target was *S. cerevisiae* Rrp47, although it was somewhat of a problematic target given that it contains a fairly disordered C-terminus, based on a DISOPRED prediction (*10*) (Fig. 4-6B). A protein that also binds to human Rrp6 and appears to have a function similar to Rrp47 was identified as C1D (but here referred to as human Rrp47) (*4*). Interestingly, the two proteins have very little sequence identity with the exception of a few speckled regions, and human Rrp47 is 43 amino acids shorter than its yeast counterpart (Fig. 4-6A). Nonetheless, human Rrp47 is well conserved in mammals—based on a ClustalW (*19*) sequence alignment—indicating the possibility of a conserved functional role in exosome mediated degradation. Although human Rrp47 also contains disordered regions (Fig. 4-6C), its shorter length could make it an easier structural target. Numerous constructs of yeast and human Rrp47 were produced in attempts at obtaining diffracting crystals of Rrp47. Protein modification methods such as *in situ* proteolysis and lysine methylation were attempted in order to produce better diffracting crystals. In addition, complexes between Rrp6 and Rrp47 were also attempted for crystallization.

4.5.1. Methods

Target Selection

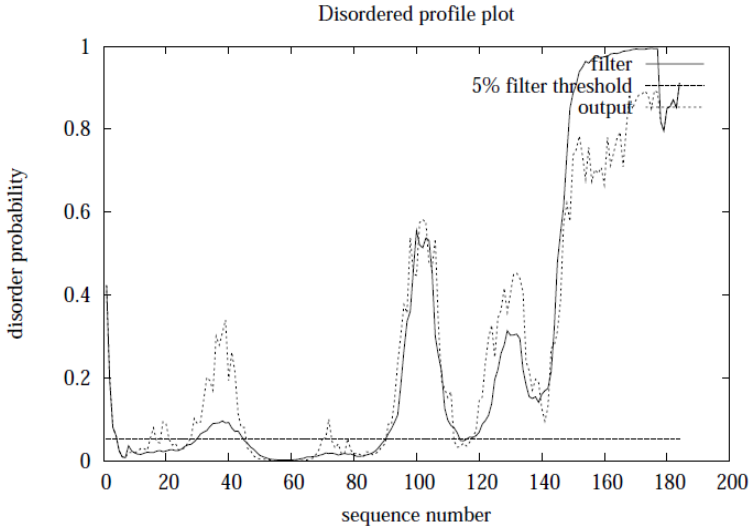
The full length versions of both the yeast and human Rrp47 were attempted, along with several truncations (Table 4-1). Truncation selection was initially based on disorder prediction (Fig. 4-6B-C). Truncations removed a varying degree of the C-terminus of Rrp47, which is fairly

Fig. 4-6. Alignment of yeast and human Rrp47, and disordered regions. A) Alignment of the two proteins. In general, there is little sequence conservation between the two proteins, with the exception of a few amino acids. Notably, yeast Rrp47 also contains two long insertions (yeast amino acids 100-127 and 150-164). Both proteins are lysine rich. The alignment was preformed with ClustalW (*19*). B) Prediction of disordered regions in yeast Rrp47 and C) human Rrp47. Disordered predictions were preformed with the DISOPRED prediction server (*10*).

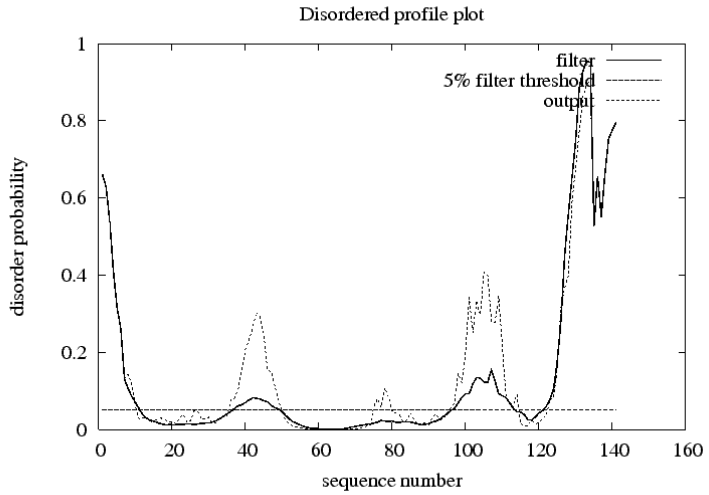
A

```
yRrp47          -------MEDIEKIKPYVRSFSKALDELKPEIEKLTSKSLDEQLLLLS-DERAKLELINR 52
hRrp47          MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELLQKLDPLEQAKVDLVS- 59
                        :   :*: *: :*.:::. :.  ::.: * * :* *  *.  *:**::*:.

yRrp47          YAYVLSSLMFANMKVLGVK-DMSPILGELKRVKSYMDKAKQYDNRITKSNEKSQAEQEKA 111
hRrp47          -AYTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAG----------- 107
                 **.*.*:::. : . **: .  *:  **:*:: **::.*:  :: . .

yRrp47          KNIISNVLDGNKNQFEPSISRSNFQGKHTKFENDELAESTTTKIIDSTDHIRKASSKKSK 171
hRrp47          ----------------KLDR----GAASRFVKNALWE---------------PKSKNAS 131
                             .:.*    *  ::* :: * *               ..**::.

yRrp47          RLDKVGKKKGGKK 184
hRrp47          KVANKGKSKS--- 141
                :: : **.*.
```

B



C



Truncation
yRrp47 FL
yRrp47 1-93
yRrp47 1-119
yRrp47 8-133
yRrp47 1-148
yRrp47 8-83
yRrp47 32-132
hRrp47 FL
hRrp47 1-123
hRrp47 8-123
hRrp47 8-end

Table 4-1. Truncations of Rrp47 for use in crystallization screening. The table lists truncations of yeast (y) or human (h) Rrp47 attempted for crystallizations.

disordered. Later truncations were based on limited proteolysis data (discussed later). Truncations of Rrp6 were based on the C-terminal boundary of the crystallized fragment (*9*) (1-536), the region shown to bind to the exosome (*11*) (1-660), and a minimal region required for binding to Rrp47 (*5*, *6*) (1-176).

Cloning and Expression

S. cerevisiae and human Rrp47, as well as all truncations were PCR amplified using primers containing a 5' end BamHI site for the yeast version, or a SacI site for the human version, and a 3' end XhoI site. Yeast genomic DNA served as a template for the *S. cerevisiae* Rrp47, and cDNA from MCF-7 cells served as the template for the human version. Inserts were cleaved with BamHI (or SacI) and XhoI and ligated into vectors containing an N-terminal hexahistidine SUMO fusion (pSUMO) and kanamycin selective marker. Plasmids were then transformed into DH5α cells and grown on LB plates supplemented with kanamycin. The full length version of Rrp6, as well as several truncations (amino acids 1-176, 1-660, and 1-536), were cloned into pET19b vectors as discussed in Section 4.4. Colonies were mini-prepped, and the presence of the correct insert was determined by DNA sequencing. Rrp47 plasmids were then transformed into BL21* cells or Rosetta II cells, either alone or co-transformed with Rrp6. Cultures were grown in 2-4 L of LB supplemented with 25 µg/mL of kanamycin and 34 µg/mL chloramphenicol (as well as 100 µg/mL ampicillin for co-transformants) and grown at 37 °C and 200 rpm until reaching an OD600 of ~1-1.7. At this point, cells were induced with 1 mM IPTG and incubated at 18 °C and 200 rpm overnight.

Protein Purification

Cells pellets were resuspended in Binding Buffer containing 50 mM sodium phosphate, 300 mM

NaCl, 10 mM Imidazole, 5 mM betamercaptoethanol, 10% glycerol, and 1 mM PMSF. Cells were lysed by sonication and pelleted to remove cell debris. Supernatant was run through a Ni-NTA column, the column was washed with 10 column volumes of Binding Buffer, and eluted with Elution Buffer (Binding Buffer containing 300 mM Imidazole). Both yeast and human Rrp47 were very salt sensitive and precipitated out of solution readily when dialyzed into low salt buffer overnight. Therefore, proteins were dialyzed into 20 mM Tris pH 8.0, 300 mM NaCl, 5 mM betamercaptoethanol, and 10% glycerol overnight at 4 °C, and the NaCl concentration was never reduced below 300 mM during further purification steps to avoid protein precipitation. During dialysis, the SUMO tag was cleaved with Sm3t SUMO protease. Rrp6/Rrp47 complexes were prepared in a similar fashion, but appeared less salt sensitive.

Yeast Rrp47 full length was re-run through a nickel column, followed by a heparin column equilibrated with dialysis buffer and eluted with a gradient of 300 mM – 2 M NaCl. The best protein purity was obtained by performing Sephacryl S-100 size exclusion chromatography in buffer containing 20 mM Tris pH 8.0, 500 mM NaCl, 10% glycerol, and 2 mM DTT. The protein was dialyzed into 20 mM HEPES pH 6.0, 300 mM NaCl, 10% glycerol, and 2 mM DTT and then run through a MonoS column in this buffer, and eluted with a gradient of 300 mM – 1 M NaCl (Fig. 4-7A). However, some batches were produced to sufficient purity without the use of the last two steps. Initially, protein was concentrated to 9-10 mg/mL for crystallization screening, although higher concentrations could be obtained in subsequent purifications.

Human Rrp47 was further purified by heparin and run on a Sephacryl S-100 size exclusion column in buffer containing 20 mM Tris pH 8.0, 300 mM NaCl, and 1 mM DTT. The protein was then concentrated in this buffer to 39.5 mg/mL. Purification of truncated forms of the yeast and human Rrp47 were attempted using methods similar to the full length versions.

Rrp6/Rrp47 complexes were further purified by heparin and eluted with a gradient of NaCl, followed by a final purification on a Superdex 200 column.

Limited Proteolysis

In order to find more stable truncations of Rrp47, limited proteolysis analysis was conducted, mainly with yeast Rrp47, as follows. The proteolysis reaction was conducted in a volume of 100 µL in buffer containing 125 mM Tris pH 8.0, 500 mM KCl, 12.5 mM magnesium acetate, and 12.5 mM $CaCl_2$. About 60 µg of Rrp47 was incubated with 60 ng of trypsin, chymotrypsin, or elastase, and incubated at room temperature. Aliquotes of 10 µL of the mixture were withdrawn at various time points, mixed with an equal volume of SDS-PAGE loading dye, and heated at 65 °C for 10 minutes to denature the protease. Samples were then run on 15% acrylamide gels and stained with colloidal coomassie blue stain to visualize protein. Bands of interest were cut out and analyzed by mass spectrometry.

Crystallization Screening

Screening was performed using nanoscale conditions with a Phoenix crystallization setup robot. Yeast Rrp47 was initially screened at 10 mg/mL, while human Rrp47 was screened at 10, 20, and 39 mg/mL.

***In-situ* proteolysis crystallization**

While work on cloning of various Rrp47 truncations commenced, crystallization screening of full length versions of yeast and human Rrp47 was conducted using the *in-situ* proteolysis technique (*20, 21*) with elastase, since it was the only promising protease from the previous *in-vitro* proteolysis experiments. Elastase was mixed with proteins in molar ratios of 1:100, 1:1,000, and 1:10,000, and incubated on ice for 30 minutes. The mixture was then set up in nanoscale crystallization trials using the JCSG+ and Proplex crystallization condition suites.

Lysine Methylation

Lysine methylation was completed based on the published protocol (*22*). Briefly, protein was diluted into 50 mM HEPES pH 7.5 and 300 mM NaCl to a final concentration of about 1 mg/mL. Amounts of 20 µL of 1 M dimethylamine-borane complex and 40 µL of 1 M formaldehyde were added per mL of the protein solution, and the solution was incubated at 4 °C for 2 hours. This process was then repeated once, followed by a final addition of 10 µL of 1 M dimethylamine-borane complex and overnight incubation at 4 °C. The protein was then purified on a Sephacryl S-100 column.

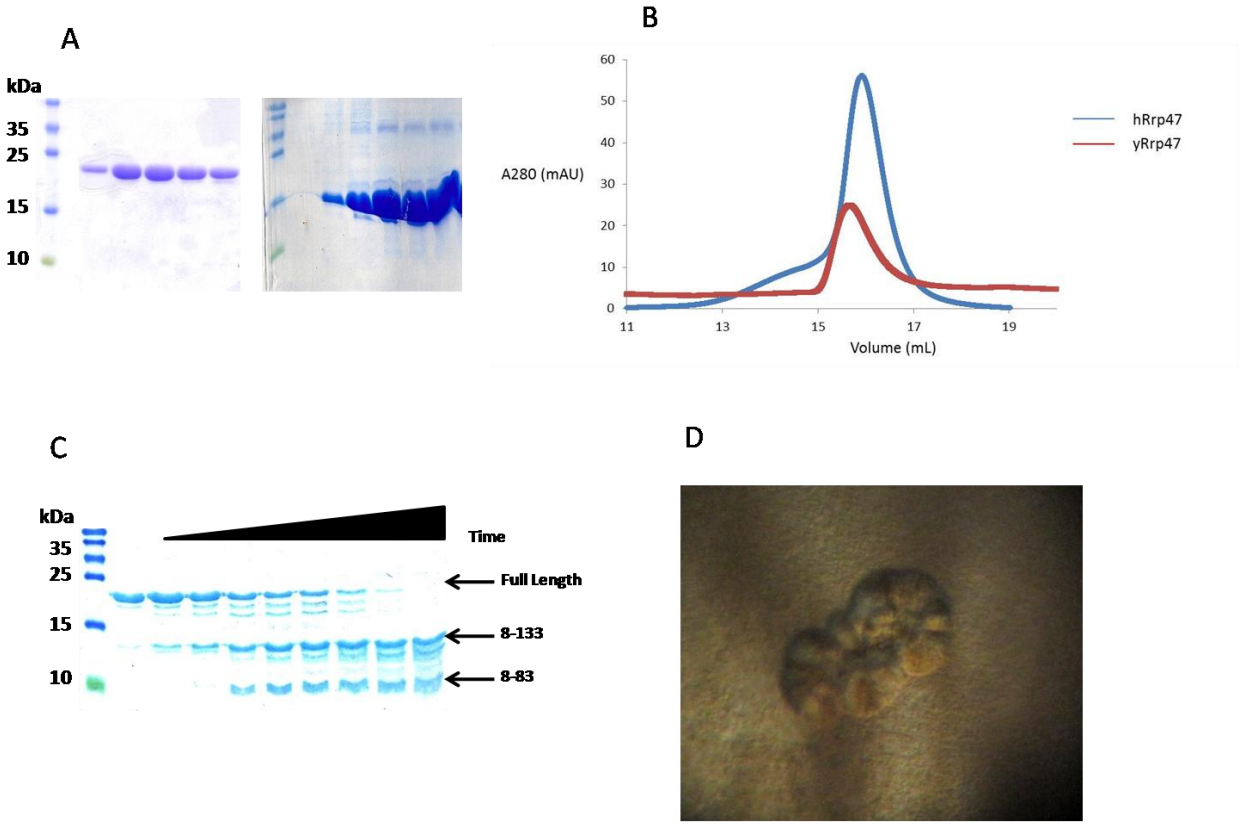
RNA degradation assay

A quick but low resolution method was used to determine whether the Rrp6 full length/Rrp47 complex retained nuclease activity. For the nuclease assays, 1.8 µM of unfolded HDV40 RNA was incubated with 1.4 µM of protein in buffer containing 25 mM Tris pH 8.0, 50 mM NaCl, 10% glycerol, and 2 mM DTT, and 5 mM of either $MgCl_2$, $MnCl_2$, or $ZnCl_2$. A time course was conducted and aliquots were removed at either 10, 30, or 60 minutes, and the reaction was stopped by addition of 2X RNA loading dye followed by heating at 65 °C for 10 minutes. Samples were then resolved through a10% acrylamide/8 M urea gel and visualized by ethidium bromide staining.

4.5.2. Results

Both full length versions of yeast and human Rrp47 could be purified to near homogeneity (Fig. 4-7A), and were soluble as long as the concentration of NaCl in the protein buffer was keep at 300 mM or higher. Notably, both proteins migrated at a position much higher than expected

Figure 4-7. Purification, proteolysis, and crystallization of Rrp47. A) Purity of yeast (left) and human (right) Rrp47. B) Size exclusion chromatography of yeast and human Rrp47 on a Superose 6 column. C) Limited proteolysis of yeast Rrp47 with elastase. D) Typical crystal clusters of Rrp47 produced from *in situ* proteolysis with elastase in crystallization conditions of 100 mM sodium acetate pH 4.6, 2.0 M ammonium sulfate.


A
kDa
35
25
15
10
B
A280 (mAU)
60
50
40
30
20
10
0
11
13
15
17
19
Volume (mL)
hRrp47
yRrp47
C
kDa
35
25
15
10
Time
Full Length
8-133
8-83
D

based on their molecular weight during size exclusion chromatography, indicating the likelihood of oligomerization (Fig. 4-7B). No crystal hits were observed for the full length versions of either protein, although many wells did contain heavy phase separation, indicating that crystallization was still a possibility.

Prior disordered prediction analysis as well as limited proteolysis with elastase identified several truncations of Rrp47, which were produced in a manner similar to the full length version. While proteolysis with trypsin and chromtrypsin did not produce any stable intermediates, proteolysis with elastase produced two bands that could represent stable intermediates (Fig. 4-7C), fragments of amino acids 8-83 and 8-133. However, for an unknown reason, the majority of truncations purified appeared to be severely degraded after SUMO cleavage. Moreover, most truncations removed disordered lysines, which subsequently altered the predicted isoelectric point of the protein from highly basic (pI ~10) to only slightly basic (pI ~8). The drastic change purification on heparin difficult, resulting in lower purity. Of the initial yeast Rrp47 proteins purified, only the truncation containing amino acids 1-93 was of sufficient quality to use for screening. However, screening with this truncation did not produce crystals, with the exception of salt crystals present in conditions containing calcium.

In situ proteolysis with elastase, however, was somewhat more successful. Yeast Rrp47 produced crystal clusters in some conditions for all ratios of protease tested, while human Rrp47 did not produce any crystals. Crystals were clusters of "disks" (Fig. 4-7D) and appeared in the following two conditions: 1) 100 mM sodium acetate pH 4.6, 2.0 M ammonium sulfate, and 2) 1.0 M $LiCl_2$, 100 mM sodium citrate pH 4.0, 20% PEG 6,000, as well as conditions that contained variations of these two precipitants. Crystals possessed a high degree of birefringence upon rotation of polarized light, and readily stained blue with Izit methylene blue solution

(Hampton Research), a protein dye. Since crystals appeared even at very low ratios of elastase (1:10,000), it was unlikely that they were protease crystals. Crystals were easily reproduced using large well hanging drop trays, and were optimized by standard methods of varying the pH of buffers and the concentrations of precipitants. While optimization could be used to produce larger crystals, the crystals still retained a disk-like or cluster shape. Interestingly, similar crystals could be produced when elastase was substituted for trypsin. The crystals were thoroughly washed, dissolved in 1 M Tris pH 8.0, and analyzed by mass spectrometry, which indicated that the crystals contained Rrp47 amino acids 32-132. The Rrp47 truncation amino acids 32-132 was cloned, expressed, and purified using a similar procedure to that of the full length version. However, this truncation also appeared to suffer from degradation during purification and did not crystallize alone. It only produced crystals with a similar morphology to the full length version when in the presence of elastase. Multiple crystals were tested for diffraction at CHESS using different cryo-protectants, but no diffraction was ever obtained.

Further optimization was performed with the full length Rrp47 in the presence of a minimal amount of elastase or trypsin. Both seeding into optimized conditions as well as into random conditions was unsuccessful at producing crystals with a different morphology. Since the protein is lysine rich, lysine methylation (*22*) was attempted in hopes of producing more well ordered crystals. In general, lysine methylated crystals grew slower than their native counterparts but only under the presence of protease and retained the same disk shape. They were not suitable for diffraction.

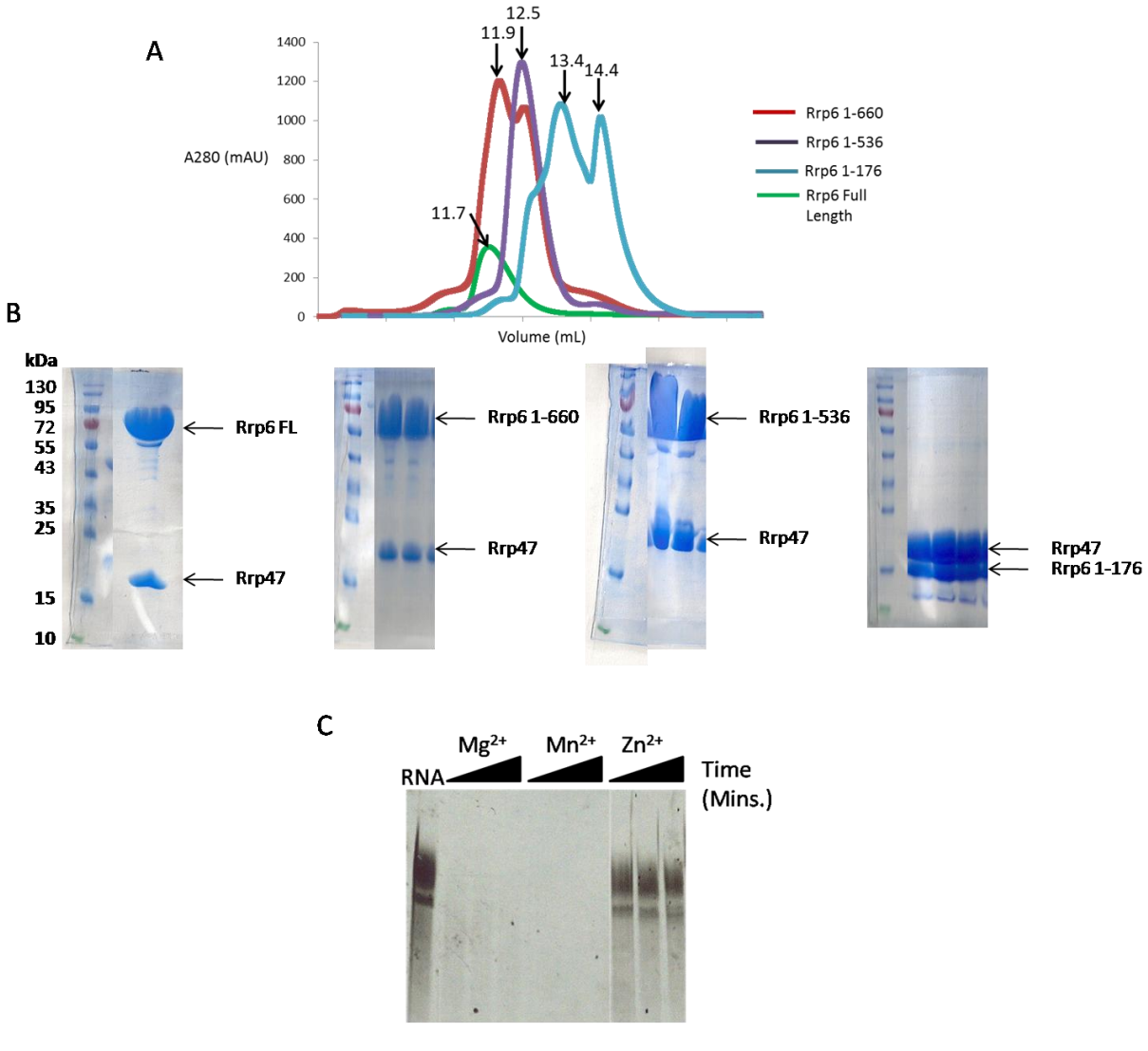
Some batches of various truncations of Rrp47 were also tested for crystallization under *in situ* proteolysis without prior removal of the N-terminal SUMO tag. While it was possible to achieve very high solubility of these proteins (upwards of 40 – 60 mg/mL concentrations), they

did not crystallize alone and produced crystals that looked identical to those produced without the tag when *in situ* proteolysis was used. These crystals were tested at CHESS, but no diffraction whatsoever was produced after testing multiple crystals in various cryo-protectants.

The Rrp6/Rrp47 complex also served as a crystallization target. All complexes eluted as a single sharp peak during heparin chromatography, suggesting the existence of a single complex. During size exclusion chromatography, the Rrp6 full length/Rrp47 complex eluted at a volume of ~12 mL, which is consistent with a molecular weight of full length Rrp6 and a hexameric Rrp47 complex (Fig. 4-8A). Complexes of Rrp47 with the truncated version of Rrp6 eluted at volumes consistent with a ratio of Rrp6 to Rrp47 of 1:6 as well (Fig. 4-8A). While the complex containing full length Rrp6 and the 1-536 produced a single peak during size exclusion chromatography, complexes containing truncations 1-660 and 1-176 produced two overlapping peaks. Therefore, it is possible that the latter complexes are composed of a varying mixture of Rrp47, although the peak doubling could also just be an effect of column over-loading (Fig. 4-8A).

To test whether the Rrp6 full length/Rrp47 complex retained nuclease activity, it was subjected to a nuclease assay with various divalent cations (magnesium, manganese, or zinc). Rrp6 displayed robust activity with either magnesium or manganese, but no activity when zinc was used (Fig. 4-8C). This indicates that the recombinantly purified Rrp6/Rrp47 retained its biological activity, and therefore is properly folded. Although all mentioned Rrp6/Rrp47 complexes were subjected to crystallization screens, none produced crystals. This suggests that remaining flexible regions on Rrp6 or Rrp47 may be impeding crystallization.

Figure 4.8. Analysis of Rrp6/Rrp47 complexes. A) Size exclusion chromatography of Rrp6/Rrp47 complexes on Superdex 200. B) SDS-PAGE analysis of complexes after three step purifications. Elution volumes of the peaks are labeled (in mL) with arrows. C) Activity assay of Rrp6 full length/Rrp47 on unfolded HDV40 ribozyme using various metal ions. Due to low resolution of this method, end products cannot be visualized.


A
12.5
11.9
13.4
14.4
11.7
A280 (mAU)
1400
1200
1000
800
600
400
200
0
Volume (mL)
Rrp6 1-660
Rrp6 1-536
Rrp6 1-176
Rrp6 Full Length
B
kDa
130
95
72
55
43
35
25
15
10
Rrp6 FL
Rrp47
Rrp6 1-660
Rrp47
Rrp6 1-536
Rrp47
Rrp47
Rrp6 1-176
C
Mg2+
Mn2+
Zn2+
RNA
Time (Mins.)

4.6. Discussion

In order to gain insight into the organization of the nuclear exosome, we have been able to reconstitute the Rrp6 containing exosome and conduct preliminary EM analysis on it. While both western blotting and 2D reconstructions indicate the presence of Rrp6 in the complex, density for Rrp6 is lost during 3D reconstruction. It is possible that Rrp6 dissociates during negatives stain preparation, indicating that Cryo-EM may be a better alternative. While this technique has been tried with the Rrp6 exosome by our collaborators, so far it has been unsuccessful (*13*). The other possibility is that Rrp6 is flexible and the density for it thus averages out. While our efforts at using cross-linking reagents to prevent flexibility were not successful, more advanced techniques could be attempted. In particular, a preparation technique called GraFix, in which a cross-linking gradient is coupled to a density gradient, could be useful (*23*). This technique allows for the selection of individual fractions from the density gradient to be used for EM analysis, allowing for the removal of aggregated fractions.

Our protein interaction analysis of the nuclear exosome revealed that accessory proteins interact with the yeast exosome in a manner different than the human exosome. While it was reported that human Mtr4 bound to both Rrp6 and Mpp6 to form a trimeric complex (*4*), our data do not support these results for *S. cerevisiae*. Instead, we detect binding between Mpp6 and the core exosome, but no interaction between Mtr4 with Rrp6, Mpp6, or the core exosome. Therefore, it is quite possible that the human proteins have diverged significantly from the yeast homologues. Moreover, the human exosome has been reported to form the larger NEXT complex (*24*) which is not known to associate with the exosome in yeast. It is, however, a bit curious that Mpp6—a nuclear protein—associates with the general core exosome instead of the

nuclear specific cofactor Rrp6. In theory, Mpp6 could also associate with the cytoplasmic exosome, but the protein has only been detected in nuclear exosome fractions. The simplest explanation is that since Mpp6 contains a nuclear localization signal it is restricted from the cytoplasm, preventing it from associating with the cytoplasmic exosome. Therefore, Mpp6 could serve to tether substrate RNA to either the exosome or Rrp6 without making protein contacts with Rrp6 itself. Further structural and functional studies would help to resolve this mechanism.

Our studies do confirm an interaction of Rrp6 with Rrp47 that is stable even under high salt conditions used during purification. We have been able to co-purify several Rrp6/Rrp47 complexes, and have produced crystals of yeast Rrp47. While it is promising that Rrp47 produced some crystals when proteolyzed by elastase, the crystals appear to be poorly ordered and do not diffract at all. Since Rrp47 has been reported to form a hexamer, it is possible that when protease is added, heterogeneous hexamers, consisting of monomers of Rrp47 truncated at different positions, are formed and crystallize. However, these crystals are likely to be poorly ordered and thus do not diffract. Since the Rrp47 fragment of amino acids 32-132, identified by mass spectrometry analysis of the crystals, only crystallized in the presence of protease as well, crystallization may be dependent on the cleavage of internal loops instead of N- or C-terminal disordered regions. Therefore, locating predicted flexible loops and systematically deleting them may aid in producing better diffracting crystals. The caveat here is that truncating Rrp47 has been shown to result in protein instability and degradation, so it is likely that many different truncations of internal loops will need to be screened in order to find one that both retains protein stability and crystallizes. Another option to improve stability is to try co-crystallizing truncated Rrp47 with Rrp6. However, this will likely require a bit of trial and error, since an Rrp47

truncation lacking only the first eight N-terminal amino acids failed to pull-down Rrp6 during co-expression (data not shown). Therefore, it is likely that more effort in protein engineering of Rrp47 (and possibly Rrp6) is needed in order to produce crystals suitable for X-ray diffraction.

References

1. Briggs, M. W., Burkard, K. T., and Butler, J. S. (1998) Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential for efficient 5.8 S rRNA 3' end formation, *The Journal of biological chemistry 273*, 13255-13263.
2. Cristodero, M., Bottcher, B., Diepholz, M., Scheffzek, K., and Clayton, C. (2008) The Leishmania tarentolae exosome: purification and structural analysis by electron microscopy, *Molecular and biochemical parasitology 159*, 24-29.
3. Lehner, B., and Sanderson, C. M. (2004) A protein interaction framework for human mRNA degradation, *Genome research 14*, 1315-1323.
4. Schilders, G., van Dijk, E., and Pruijn, G. J. (2007) C1D and hMtr4p associate with the human exosome subunit PM/Scl-100 and are involved in pre-rRNA processing, *Nucleic acids research 35*, 2564-2572.
5. Stead, J. A., Costello, J. L., Livingstone, M. J., and Mitchell, P. (2007) The PMC2NT domain of the catalytic exosome subunit Rrp6p provides the interface for binding with its cofactor Rrp47p, a nucleic acid-binding protein, *Nucleic acids research 35*, 5556-5567.
6. Costello, J. L., Stead, J. A., Feigenbutz, M., Jones, R. M., and Mitchell, P. (2011) The C-terminal region of the exosome-associated protein Rrp47 is specifically required for box C/D small nucleolar RNA 3'-maturation, *The Journal of biological chemistry 286*, 4535-4543.
7. Schilders, G., Raijmakers, R., Raats, J. M., and Pruijn, G. J. (2005) MPP6 is an exosome-associated RNA-binding protein involved in 5.8S rRNA maturation, *Nucleic acids research 33*, 6795-6804.
8. Milligan, L., Decourty, L., Saveanu, C., Rappsilber, J., Ceulemans, H., Jacquier, A., and Tollervey, D. (2008) A yeast exosome cofactor, Mpp6, functions in RNA surveillance and in the degradation of noncoding RNA transcripts, *Molecular and cellular biology 28*, 5446-5457.
9. Midtgaard, S. F., Assenholt, J., Jonstrup, A. T., Van, L. B., Jensen, T. H., and Brodersen, D. E. (2006) Structure of the nuclear exosome component Rrp6p reveals an interplay between the active site and the HRDC domain, *Proceedings of the National Academy of Sciences of the United States of America 103*, 11898-11903.
10. Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F., and Jones, D. T. (2004) The DISOPRED server for the prediction of protein disorder, *Bioinformatics 20*, 2138-2139.
11. Callahan, K. P., and Butler, J. S. (2008) Evidence for core exosome independent function of the nuclear exoribonuclease Rrp6p, *Nucleic acids research 36*, 6645-6655.
12. Wang, H. W., Wang, J., Ding, F., Callahan, K., Bratkowski, M. A., Butler, J. S., Nogales, E., and Ke, A. (2007) Architecture of the yeast Rrp44 exosome complex suggests routes of RNA recruitment for 3' end processing, *Proceedings of the National Academy of Sciences of the United States of America 104*, 16844-16849.
13. Liu, X., and Wang, H. W. (2010) Preliminary Electron Microscopy Studies of Recombinant Exosomes, Yale University School of Medicine.
14. Ludtke, S. J., Baldwin, P. R., and Chiu, W. (1999) EMAN: semiautomated software for high-resolution single-particle reconstructions, *Journal of structural biology 128*, 82-97.
15. Greimann, J. C., and Lima, C. D. (2008) Reconstitution of RNA exosomes from human and Saccharomyces cerevisiae cloning, expression, purification, and activity assays, *Methods in enzymology 448*, 185-210.

16. Mitchell, P., Petfalski, E., Houalla, R., Podtelejnikov, A., Mann, M., and Tollervey, D. (2003) Rrp47p is an exosome-associated protein required for the 3' processing of stable RNAs, *Molecular and cellular biology 23*, 6982-6992.
17. Mitchell, P. (2010) Rrp47 and the function of the Sas10/C1D domain, *Biochemical Society transactions 38*, 1088-1092.
18. Yavuzer, U., Smith, G. C., Bliss, T., Werner, D., and Jackson, S. P. (1998) DNA end-independent activation of DNA-PK mediated via association with the DNA-binding protein C1D, *Genes & development 12*, 2188-2199.
19. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic acids research 22*, 4673-4680.
20. Dong, A., Xu, X., Edwards, A. M., Chang, C., Chruszcz, M., Cuff, M., Cymborowski, M., Di Leo, R., Egorova, O., Evdokimova, E., Filippova, E., Gu, J., Guthrie, J., Ignatchenko, A., Joachimiak, A., Klostermann, N., Kim, Y., Korniyenko, Y., Minor, W., Que, Q., Savchenko, A., Skarina, T., Tan, K., Yakunin, A., Yee, A., Yim, V., Zhang, R., Zheng, H., Akutsu, M., Arrowsmith, C., Avvakumov, G. V., Bochkarev, A., Dahlgren, L. G., Dhe-Paganon, S., Dimov, S., Dombrovski, L., Finerty, P., Jr., Flodin, S., Flores, A., Graslund, S., Hammerstrom, M., Herman, M. D., Hong, B. S., Hui, R., Johansson, I., Liu, Y., Nilsson, M., Nedyalkova, L., Nordlund, P., Nyman, T., Min, J., Ouyang, H., Park, H. W., Qi, C., Rabeh, W., Shen, L., Shen, Y., Sukumard, D., Tempel, W., Tong, Y., Tresagues, L., Vedadi, M., Walker, J. R., Weigelt, J., Welin, M., Wu, H., Xiao, T., Zeng, H., and Zhu, H. (2007) In situ proteolysis for protein crystallization and structure determination, *Nature methods 4*, 1019-1021.
21. Wernimont, A., and Edwards, A. (2009) In situ proteolysis to generate crystals for structure determination: an update, *PloS one 4*, e5094.
22. Walter, T. S., Meier, C., Assenberg, R., Au, K. F., Ren, J., Verma, A., Nettleship, J. E., Owens, R. J., Stuart, D. I., and Grimes, J. M. (2006) Lysine methylation as a routine rescue strategy for protein crystallization, *Structure 14*, 1617-1622.
23. Kastner, B., Fischer, N., Golas, M. M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., Hauer, F., Wolf, E., Uchtenhagen, H., Urlaub, H., Herzog, F., Peters, J. M., Poerschke, D., Luhrmann, R., and Stark, H. (2008) GraFix: sample preparation for single-particle electron cryomicroscopy, *Nature methods 5*, 53-55.
24. Lubas, M., Christensen, M. S., Kristiansen, M. S., Domanski, M., Falkenby, L. G., Lykke-Andersen, S., Andersen, J. S., Dziembowski, A., and Jensen, T. H. (2011) Interaction profiling identifies the human nuclear exosome targeting complex, *Molecular cell 43*, 624-637.

Appendix

Studies of Ski7: a GTPase that associates with the cytoplasmic exosome and functions in mRNA surveillance

Abstract

This appendix details the purification and crystallization attempts of Ski7, a cytoplasmic exosome associated protein. Although the full length version of the protein appeared toxic to *E. coli* cells, we were able to purify the C-terminal region of the protein (amino acids) to sufficient yield for crystallization screening.

A1.1. Introduction

Ski7 is a protein that has been reported to bridge interactions between the cytoplasmic exosome and the SKI complex to function in mRNA surveillance. Since it has sequence homology to the previously crystallized regions of *S. cerevisiae* EF1A and the C-terminal region of *S. pombe* eRF3, this homology region (Ski7 amino acids 250-747) served as the primary structural target. EF1A is involved in bringing aminoacylated tRNA to the ribosome and eRF3 is involved in peptide release after translation and has a GTPase activity that is stimulated by association with the ribosome and the eRF1 protein cofactor. Theories speculate that Ski7 may also function in peptide release during mRNA surveillance, given that its C-terminus is homologous to eRF3. We attempted a structural approach to better characterize the role of Ski7. The following strategies were used in attempts to purify and crystallize the C-terminal region of Ski7.

A1.2. Methods

Cloning

Previous studies indicated that Ski7 did not express in typical expression vectors, possibly due to toxicity of the protein to *E. coli* (Ke lab, unpublished data). Therefore, Ski7 was instead cloned into a pBAD vector, which contains more stringent control over expression by the inducer arabinose. Ski7 full length as well as truncations (1-255, 200 – 747, 260 – 747, and 250 – 747) were PCR amplified using a 5' primer containing a NcoI site and N-terminal hexahistidine region, and a 3' primer with a XhoI site and inserted into a pBAD vector linearized with NcoI and XhoI restriction enzymes. DNA was sequenced to confirm the correct insert. The proteins were initially purified by Ni-NTA chromatography, followed by ion exchange. While the full length and N-terminal (1-255) truncations did not expressed, C-terminal truncations expressed to a certain extent. The best protein quality and yield was obtained for Ski7 250 – 747. The following protocol is an optimized method for preparing batches of Ski7 250 – 747 of sufficient purity and quantity for crystallization attempts.

Expression and Purification

A volume of 10 L of LB was inoculated with 1 mL of Ski7 250 – 747 overnight culture/L and supplemented with 100 µg/mL of ampicillin. Cultures were grown at 37 °C until reaching an OD600 of 1.0. Cultures were then induced with 10 mL of 20% arabinose at 37 °C and induced for ~8 hours at 37 °C. Cultures were induced at 37 °C instead of 4 °C because low temperature induction resulted in a high migrating contaminant after SDS-PAGE analysis that was suspected to be a chaperone that could have resulted from cold shock to the bacteria. Cells were pelleted, snap frozen, and stored at -80 °C until use.

For purification, cells were thawed and resupsended in a total volume of 150 mL with Buffer A (50 mM sodium phosphate pH 8.0, 300 mM NaCl, 10 mM imidazole, 10% glycerol, 5 mM betamercaptoethanol, 10 µg/mL DNaseI, and 1X protease inhibitors). Cells were lysed by sonication and centrifuged to pellet cell debris. Supernatant was then loaded onto Ni-NTA resin equilibrated in Buffer A. Resin was then washed with 50 mL of Buffer A, followed by 30 mL of Buffer A containing 20 mM imidazole. Protein was eluted with Buffer A containing 250 mM imidazole. Imidazole was removed and NaCl reduced by dialysis into 20 mM Tris pH 8.0, 50 mM NaCl, 10% glycerol, and 5 mM betamercaptoethanol overnight at 4 °C.

Since Ski7 250 – 747 had previously been shown to precipitate on a MonoQ column during loading, the protein was instead subjected to a Q fast flow column (GE Healthcare), and loaded in three rounds with a reduced protein concentration per round. After this step, protein was mostly pure, with the exception of a contaminant that ran at a molecular weight of ~75 kDa. This contaminant was effectively removed by concentrating the protein and running it through size exclusion chromatography on a Sephacryl S-100 column, equilibrated in 20 mM Tris pH 8.0, 2 M NaCl, 10 % glycerol, and 2 mM DTT. After size exclusion chromatography, the protein was judged to be ~99% pure by SDS-PAGE (Fig. A-1). Peak fractions were pooled and dialyzed into 20 mM Tris pH 8.0, 200 mM NaCl, and 2 mM DTT, which was the final buffer for crystallization screening. Ski7 truncations 200-end and 260-end were prepared in a similar fashion, but their final yields were too low for crystallization screening.

Initial Crystallization screening

Ski7 250 – end was concentrated to 18.2 mg/mL and initially set up in crystallization trays at concentration of 12 and 8 mg/mL at room temperature and at 4 °C. Since Ski7 has 250 – 747 is

homologous to GTPase proteins, crystallization was also attempted in the presence of 2 mM $MgCl_2$ and either GTP or GDP at both room temperature and 4 °C.

Optimal Solubility Method Screening

One issue with the protein was that it precipitated out of solution readily when it was let sit for ~30 minutes on ice at concentrations of 5 – 8 mg/mL. It was therefore thought that the protein may be poorly soluble in the buffer in which it was concentrated (20 mM Tris pH 8.0, 200 mM NaCl, and 2 mM DTT). Screening using the optimal solubility (OS) method was attempted according to the published protocol (*1*). Briefly, 1 µL of protein was mixed with 1 µL of various buffers at 100 mM concentration at different pH, and incubated in a hanging drop format over a reservoir of the same buffer.

A1.3. Results

None of the initial crystallization screens produced crystals in either the presence or absence of 2 mM $MgCl_2$, GTP, or GDP at room temperature or 4 °C incubation. Most conditions produced heavy precipitation. Of 24 buffers tested for the OS method, only three did not produce precipitation after overnight incubations: 100 mM solutions of glycine at pH 3.0 and CHES at pH 9.0, and 9.5. Since extremely low pH solutions can denature proteins, CHES at pH 9.0 was chosen as a good buffer to use to help to keep the protein soluble. Another batch of Ski7 250 – 747 was purified using a similar protocol to that described, but the final size exclusion step was instead preformed in 20 mM CHES pH 9.0, 150 mM NaCl, and 2 mM DTT. The protein was concentrated to ~11 mg/mL and screened for crystallization. While the solubility of the protein was greatly improved, no crystals were produced and many conditions remained clear.

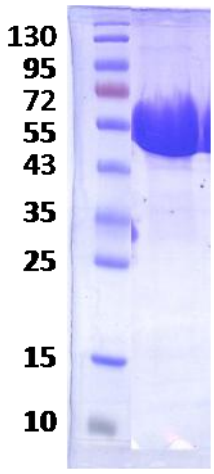


Figure A-1. SDS-PAGE analysis of Ski7 250-747 after three purification steps.

A1.4. Discussion

Although the yield of Ski7 during purification is low, we were able to obtain a large amount of protein suitable for crystallography studies of a truncation comprising amino acids 250 – 747, and smaller batches of other truncations (260 – 747 and 200 – 747). However, precipitation occurred readily in Ski7 250 – 747 at protein concentrations of 6 – 8 mg/mL or higher. Although the protein was more soluble in mM CHES pH 9.0, 150 mM NaCl, and 2 mM DTT, no crystals were produced during screening in this buffer either.

We were not able to purify either the full length version or the N-terminal region of Ski7 which limited our studies of the Ski7 bound exosome. Although the C-terminal truncations of Ski7 were tested for their ability to bind to the exosome core, no stable association was established (data not shown), which agrees with previous results that suggest that the N-terminus of Ski7 appears to be the major site of attachment to the exosome (*2*). A batch of Ski7 truncations 200 – 747 and 260 – 747 was tested for GTPase activity and shown to be an active GTPase by Chris Shoemaker of Johns Hopkins (personal communication). Moreover, this activity was not stimulated by either the ribosome or other cofactors, and we were not able to find a stable association between Ski7 and either amino acylated tRNA or the ribosome (Chris Shoemaker, personal communication). Therefore, as far as we know, Ski7 does not appear to function in a similar manner to either EF1A or eRF3. However, further studies will need to be conducted to determine the exact role of Ski7 in exosome mediated degradation.

References

1. Jancarik, J., Pufan, R., Hong, C., Kim, S. H., and Kim, R. (2004) Optimum solubility (OS) screening: an efficient method to optimize buffer conditions for homogeneity and crystallization of proteins, *Acta crystallographica. Section D, Biological crystallography 60*, 1670-1673.
2. Araki, Y., Takahashi, S., Kobayashi, T., Kajiho, H., Hoshino, S., and Katada, T. (2001) Ski7p G protein interacts with the exosome and the Ski complex for 3'-to-5' mRNA decay in yeast, *The EMBO journal 20*, 4684-4693.